Aus der Direktion Medizin, Inselspital, Inselgruppe
Direktor: Prof. Dr. Martin Fiedler

Arbeit unter der Leitung von:
Prof. Dr. phil. Rouven Porz

Ethische Überlegungen zur Künstlichen Intelligenz in der Medizin

Eine qualitative Interviewstudie als Standortbestimmung

Inaugural-Dissertation zur Erlangung der Doktorwürde der Humanmedizin
der Medizinischen Fakultät der Universität Bern

vorgelegt von:

MARIA KATHARINA MAHLER
von Uzwil SG

Von der Medizinischen Fakultät der Universität Bern auf Antrag der Dissertationskommission als Dissertation genehmigt.

Promotionsdatum:

Der Dekan der Medizinischen Fakultät:

Vorwort

Kann eine Maschine denken lernen? Bereits während des Studiums weckte das Thema Künstliche Intelligenz und deren Potenzial grosse Faszination in mir. Je mehr ich mich mit dem Thema beschäftigte, desto mehr Fragen stellten sich mir. Ist es möglich, dass eine KI Bewusstsein entwickelt? Und woran könnte man das merken? Was passiert, wenn eine Maschine intelligenter wird als ein Mensch? Wo stehen wir in der Medizin in Bezug auf die Anwendung von KI?

Ich sehe jeden Tag, an dem ich im Spital arbeite, Möglichkeiten zum Einsatz von KI. Es gäbe vieles, was den Arbeitsalltag erleichtern könnte - intelligente Tools könnten gewisse Arbeitsschritte beschleunigen und vereinfachen. Beispiele mit viel Potential sind Dokumentationsprogramme und Patientenmanagementsysteme. KI könnte helfen, uns Ärztinnen und Ärzte zu entlasten; uns mehr Zeit für das zu geben, was im Zentrum der Medizin stehen sollte: Der Mensch.

Die Anwendung von KI in der Medizin bringt ethische Fragen mit sich.

1. Wer trägt die Verantwortung, wenn eine KI einen Fehler macht?
2. Soll eine KI Entscheidungen über Leben und Tod fällen dürfen?
3. Wird KI die soziale Ungerechtigkeit in der Medizin verstärken?
4. Ist Empathie programmierbar?
5. Wären Sie bereit, alle Ihre Daten (alles, was Sie denken, fühlen, machen, sehen etc.) preiszugeben für eine hundertprozentig genaue Diagnose?

Aus dem Wunsch heraus, solchen Fragen auf den Grund zu gehen, entstand die vorliegende Arbeit. Zusammen mit meinem Betreuer, dem Medizinethiker Prof. Dr. Rouven Porz, habe ich mich dazu entschieden, diese Fragen Expertinnen und Experten aus dem Bereich der medizinischen KI-Entwicklung zu stellen. Sind sie sich der Tragweite ihrer Arbeit bewusst? Welche ethischen Herausforderungen sehen sie in Zukunft?

Eines ist klar: KI wird immer mehr in den medizinischen Alltag integriert werden. Die Frage ist nur, wie wir mit ihr umgehen.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Prometheus modellierte aus Ton den ersten Menschen nach dem Ebenbild der Götter. Athena, die Göttin der Weisheit, hauchte ihm das Leben ein. In «Frankenstein» (1818) von Mary Shelley (1797 – 1851), auch genannt «Der moderne Prometheus» wird ein lebendiges Wesen aus Leichenteilen erschaffen. Beide Geschichten, jene der griechischen Mythologie und diese von Frankenstein, behandeln die Erschaffung von Leben. Die Schöpfungen werden sowohl von den Göttern in der Sage von Prometheus als auch von den Menschen in der Erzählung von Mary Shelley mit Ablehnung, Kritik und Missgunst beantwortet.[1] Prometheus wird von Zeus in den Kaukasus verbannt. Jeden Tag wird er von einem Adler heimgesucht, der sich an seiner Leber nährt, die immer wieder aufs Neue regeneriert. In Frankenstein entwickelt das Wesen einen Groll gegenüber den Menschen aufgrund der Ablehnung, die er durch sie erfährt.[2] Die Geschichte Frankensteins ist noch immer sehr aktuell. Sie widerspiegelt in sehr guter Art und Weise die Ängste und Bedenken der Gesellschaft gegenüber künstlich erschaffenen Technologien. Die Befürchtung, dass sich das künstlich erschaffene intelligente Wesen zu etwas Bösem und Unkontrollierbarem entwickeln könnte, scheint nicht neu zu sein.

Nicht immer hat sich der Traum der Schöpfung neuen Lebens erfüllt. Michelangelo verbrachte sein ganzes Leben damit, Ebenbilder des Menschen aus Stein nachzubilden. Nach jahrelanger Arbeit vollendete er seine Marmorskulptur «Mosè». Im Gegensatz zu den Geschichten von Prometheus und Frankenstein gelang es ihm nicht, der Schöpfung das Leben einzuhauchen. Enttäuscht vom leblosen Mosè habe er gemäss einer Legende mit dem Hammer auf dessen Knie geschlagen und folgendes gesagt:

> *Perché non parli?*
>
> Deutsch so viel wie: «Warum sprichst du nicht?»

Diese Erzählungen zeigen sehr gut, dass seit jeher Schöpfungen, die der Intelligenz und dem Bewusstsein des Menschen sehr nahekommen, Faszination und Furcht auslösen.

Im sozialutopischen Drama des tschechischen Schriftstellers Karel Čapek (1890 – 1938) mit dem Namen «R.U.R. - Rossum's Universal Robots» bzw. tschechisch «Rossumovi Univerzální Roboti» findet sich dieselbe Thematik. Das Drama handelt von künstlichen Menschen «robots», die kreiert wurden, um menschliche Arbeiten zu verrichten. Das Wort Roboter hat seinen Ursprung in dieser Geschichte von Čapek. Es leitet sich aus dem tschechischen «robota» ab, was mit Zwangsarbeit übersetzt werden kann.[3]

In kaum einem Gebiet verschwimmen Realität und Science Fiction derart wie bei Künstlicher Intelligenz. Popkulturelle Filme haben unser Bild von Künstlicher Intelligenz (KI) geprägt und auch unsere Bedenken und Ängste gegenüber diesen beeinflusst. Filme wie «2001 a Space Odyssey (1968)», «Ex Machina (2015)», «The Matrix (1999)», «TAU (2018)» oder die Serie «Black Mirror (2011)» haben einen Effekt darauf, wie wir die Verwendung von KI und deren Auswirkungen auf die Gesellschaft wahrnehmen. Im alltäglichen Leben begegnen wir der KI in Gestalt von Sprachassistenten wie Siri (Apple) oder Alexa (Amazon), autonom fahrenden Autos

von Tesla oder Google, personalisierter Werbung auf Facebook, Algorithmen zur Erkennung von Vorlieben auf Instagram, Netflix und YouTube oder auch Chatbots wie ChatGPT von OpenAI. Bei einer kritischen Betrachtung von vorherig genannten, gängigen Anwendungsbeispielen von KI stellen sich schnell Fragen, wie die folgenden:

- Wer trägt die Verantwortung, wenn ein selbstfahrendes Auto einen Unfall verursacht?
- Kann ich ChatGPT vertrauen?
- Wie kennt Netflix meine Vorlieben?
- Kann ein Sexroboter Liebe empfinden?
- Darf ein Kampfroboter eigenmächtig entscheiden und einen Menschen töten?

Was haben diese Fragen mit Medizin zu tun? Zunächst scheint es, als hätten die Fragestellungen keine unmittelbare Beziehung zur Medizin. Doch auch in der Medizin gibt es immer mehr KI-Programme. Vermutlich werden Programme, Sprachassistenten und andere intelligente Algorithmen, die wir aus dem Alltag kennen, bald auch in der Medizin zunehmend zur Anwendung kommen. Die Anwendung von KI bringt viele Vorteile mit sich: Sie könnte zu einer Zeitersparnis von vielen ärztlichen und administrativen Tätigkeiten führen, akkuratere Diagnosen stellen, zu einer Reduktion der Arbeitslast der Ärzte beitragen oder eine Kostensenkung im Gesundheitssystem als Effekt haben. KI hat das Potenzial, das Gesundheitswesen von Grund auf zu verändern. Gerade hier stellen sich jedoch brisante ethische Fragen, die sich kaum von anderen Anwendungsgebieten unterscheiden. Das zeigen folgende zwei Beispiele:

- Wer trägt die Verantwortung, wenn eine diagnostische KI eine fehlerhafte Diagnose erstellt?
- Muss ein Pflegeroboter empathische Fähigkeiten haben?

Die kommenden Seiten gehen auf die Themengebiete Ethik, Medizinethik und Technologieethik sowie auf Künstliche Intelligenz und KI in der Medizin ein. Es soll zuerst ein Überblick über das Themengebiet geschaffen werden und das nötige Grundlagenwissen zur Verfügung gestellt werden, damit die Forschungsfrage und die Ergebnisse der Interviews mit den Expertinnen und Experten aus dem Bereich der KI-Entwicklung nachvollzogen werden können. Die Einleitung soll den aktuellen Forschungs- und Wissensstand möglichst gut beleuchten. Zu Beginn werden die Themengebiete einzeln erläutert und dann später miteinander in Verbindung gebracht. Die erläuterten Themen sind in der Forschungsfrage vereint.

1.1 Ethik

Der Begriff Ethik kommt vom altgriechischen Wort «êthikê», was Sitte, Gewohnheit oder Brauch bedeutet. Ethik ist eine Disziplin der Philosophie und kann folgendermassen beschrieben werden: «Die Grundlagen der Ethik betreffen das Gute, das Haltung und Handeln des Menschen bestimmen soll. Ihr Ziel ist es, methodisch gesichert, die Grundlagen für gerechtes, vernünftiges und sinnvolles Handeln und (Zusammen-)Leben aufzuzeigen.»[4]

Der Ursprung der Ethik als Teilbereich der (westlichen) Philosophie ist im antiken Griechenland begründet. Aristoteles beschäftigte sich bereits im 4. Jahrhundert v.Chr. in seinem Werk «Nikomachische Ethik» mit der herkömmlichen Moral, wie sie sich in Sitten und Bräuchen zeigte. Ebenso beschäftigte er sich darin mit der Frage nach einem gelungenen und guten Leben und welches das höchste Gut darin ist.[5]

Ethik ist ein sehr weit gefasster Begriff, der darüber hinausgeht, was richtig und was falsch ist. Die Ethik beschäftigt sich mit den Fragen: Wie soll ich handeln? Was ist moralisch richtig? Was ist gerecht und was ist ungerecht? Was soll erlaubt sein? Was ist Moral? Die Ethik versucht, diese Fragen zu stellen, zu beantworten und zu begründen. In der Philosophie wird zwischen Ethik und Moral (aus dem Lateinischen «die Sitte betreffend») unterschieden. Als Moral werden die Sitten und Normen in einer Gesellschaft oder Gemeinschaft bezeichnet. Sie zeigt das moralisch Richtige und Falsche an. Umgangssprachlich findet jedoch meist eine Gleichstellung der Begriffe Moral und Ethik statt. In den Interviews dieser Arbeit wurden diese beiden Begriffe sehr häufig synonym verwendet.

1.2 Medizinethik

In der Medizinethik wird meistens auf die vier Prinzipien nach Beauchamp und Childress [6] Bezug genommen. Sie befasst sich sowohl mit der Anwendung dieser Prinzipien im Gesundheitswesen als auch mit dem Handeln der Akteure in der Medizin.

- Respekt vor der Autonomie (respect of autonomy)
- Prinzip des Nichtschadens (nonmaleficence)
- Prinzip der Fürsorge (beneficence)
- Gerechtigkeit (justice)

Diese wurden als Grundprinzipien der Bioethik entwickelt und dienen als Grundlage einer ethischen Ausübung des medizinischen Berufs. Sie helfen, Lösungen ethischer Probleme und Dilemmata zu erarbeiten. Sie dienen als Rahmen und nicht als strikte Richtlinien und sind in *Tabelle* 1 erläutert.

Respekt vor der Autonomie (respect of autonomy)	**Prinzip des Nichtschadens (nonmaleficence)**	**Prinzip der Fürsorge (beneficence)**	**Gerechtigkeit (justice)**
Der Patient kann über seine Gesundheit eigenverantwortlich entscheiden. Dies geschieht in der Praxis mit Hilfe der informierten Einwilligung (engl. informed consent) um einen ärztlichen Paternalismus zu verhindern. Der Arzt oder die Ärztin muss den Patienten ausreichend informieren, damit dieser eigenmächtig entscheiden kann.	Der zweite Grundsatz beruht auf dem Prinzip «primum non nocere»; Zuerst nicht schaden. Behandlungen und Interventionen sollen nicht zu Schaden führen.	Dieses Prinzip bedeutet die Verpflichtung den Patienten mit Krankheiten zu behandeln oder diese präventiv zu verhindern. Bei gewissen Massnahmen müssen Risiko (Schaden) und Nutzen genau abgewogen werden.	Medizin und Gesundheitsressourcen sollen gerecht und fair verteilt sein. Dieses Prinzip stellt den Anspruch, alle Patienten unabhängig von ihrem biologischen Geschlecht, ihrem Gender, ihrem sozioökonomischen Status, ihrer Herkunft oder anderer Eigenschaften, gerecht zu behandeln.

Tabelle 1: Prinzipien der Medizinethik

In der Ethik zu unterscheiden sind *normative* und *evaluative* Fragen des moralisch Richtigen. Normative Fragen ergründen das moralisch Richtige und Falsche. Evaluative Fragen befassen sich hingegen mit dem guten und gelingenden Leben. Diese zwei Bereiche können unterschiedlich zum Tragen kommen. Normativ wäre beispielsweise die Achtung der Patientenautonomie. Der Patient muss über seine Behandlung informiert werden und trifft seine Entscheidung selbständig. Eine evaluative Fragestellung wäre, ob ein Leben noch lebenswert ist, wenn ein Patient ein Locked-In-Syndrom[1] erleidet.

Ein der Medizin angegliedertes Teilgebiet ist die **Public-Health-Ethik**. Diese setzt sich mit ethischen Fragen der öffentlichen Gesundheit auseinander. In der Medizinethik steht die Arzt-Patienten-Beziehung im Zentrum, in der Public-Health-Ethik die Beziehung zwischen staatlichen und nichtstaatlichen Institutionen und der Gesellschaft. Aufgaben können beispielsweise Prävention (Impfungen oder Rauchverbote) oder Screening-Programme sein.

Public-Health-Ethik bedient sich den Theorien des *Utilitarismus*. Der moderne Utilitarismus geht hauptsächlich auf den Philosophen John Stuart Mill (1806-1873) zurück, der die Ideen von Jeremy Bentham (1748-1832) weiterentwickelt und Theorien dazu definiert hat. Ziel des Utilitarismus ist das grösstmögliche Glück für die grösstmögliche

[1] Das Locked-In-Syndrom ist ein Leiden, bei dem eine komplette Lähmung der Muskulatur und der Hirnnerven (in der Regel infolge eines Schlaganfalls)vorliegt. Somit ist praktisch jegliche Kommunikation nach aussen unterbunden.

Anzahl von Menschen zu erreichen. Als moralische Pflicht des Utilitarismus gilt die Förderung des Glücks und die Minderung des Leids und des Schmerzes. Für Mill waren nicht nur die Maximierung und Minimierung dieser Zustände wichtig, sondern auch die Qualität derselben. Nicht alles Leid oder Glück ist gleich. Es komme ebenso auf die Art und Weise an, wie diese in Erscheinung treten, denn nicht alle Zustände seien gleich wertvoll.[4,5]

Bei der Abwägung von Kosten und Nutzen für die Gesellschaft sowie die Individuen wird auf die Theorien des Utilitarismus zurückgegriffen. Eine Betrachtungsweise wie diese birgt jedoch die Gefahr, dass lediglich das Wohlergehen der Mehrheit betrachtet wird. Ebenso wichtig für die Public-Health-Ethik ist deshalb die Allgemeine Erklärung der Menschenrechte, welche die UN-Generalversammlung 1948 verabschiedet hat. Darin wird der Grundsatz von Gleichheit und Freiheit festgehalten, wie auch das Recht auf Leben und Sicherheit, der Anspruch auf Rechtsgleichheit, der Schutz der Privatsphäre oder auch das Verbot der Diskriminierung. Alle diese Aspekte spielen in der Medizin eine grosse Rolle und müssen beachtet werden.[7]

Egger, Razum und Rieder [7] beschrieben die Prinzipien der Public-Health-Ethik folgendermassen:

- Gegenseitige Abhängigkeit (Interdependence): Eine Person ist nicht nur für sich selbst verantwortlich. Jede Aktion einer Person beeinflusst auch andere Personen.

- Mitwirkung (Participation): Alle haben ein Mitspracherecht, Massnahmen für die öffentliche Gesundheit sollen mit dem Einverständnis der Bevölkerung geplant werden.

- Wissenschaftliche Abstützung (Scientific evidence): Entscheide und Massnahmen müssen evidenzbasiert und wissenschaftlich abgestützt sein.

In gewissen Punkten müssen die Prinzipien der Medizinethik (die nur das Individuum betrachten) relativiert werden. Für das Wohl der Bevölkerung kann beispielsweise das Prinzip der Autonomie eines Einzelnen eingeschränkt werden. Ein Beispiel dafür sind Sicherheitsgurte. Sie erhöhen die Sicherheit beim Autofahren, können jedoch für einzelne Personen einschränkend sein.

1.3 Technologieethik und ihre Anwendung auf KI in der Medizin

Aufgrund der zunehmenden Spezialisierung und Technologisierung der Medizin entstehen aktuell immer mehr neue Bereichsethiken, da es neue Grundsätze für eine ethische Medizin braucht. Die Technologieethik ist ein sehr neues Gebiet. Sie zeigt ethische Überlegungen zur Entwicklung und Anwendung von KI auf, sowohl generell als auch in der Medizin. Durch den rasanten Fortschritt der Technik, auch in der Medizin, stehen uns plötzlich neue Möglichkeiten offen. Pflegeroboter, Gehirnimplantate, Diagnose-Apps, DNA-Analysen mittels neurologischer Netzwerke, Erkennung von Lungenkrankheiten mithilfe von Audioaufnahmen oder die automatische Analyse von

radiologischen Bildaufnahmen - dies sind nur wenige Beispiele für neue Technologien, die momentan entwickelt werden oder bereits auf dem Markt sind.

Welchen Einfluss werden diese Technologien auf den Menschen und die Medizin haben? Werden unsere Wertvorstellungen und moralischen Ansichten durch diese verändert? Wie sollen wir uns gegenüber diesen Technologien verhalten? Müssen wir uns von diesen bedroht fühlen? Kann ein Roboter Empathie empfinden? Kann eine KI moralisch korrekt handeln?

Die Technologieethik befasst sich mit einer Vielzahl an Fragen. Hier geht es nicht nur darum, was erlaubt sein soll und was nicht. Oder wie Nyholm [8] in seinem Buch «This is Technology Ethics» sagt:

> *«In a way, technology ethics can be likened to the art of jazz music. A lot of jazz is based on a set of standard compositions, which provide the basic structure of the music and main melodies and motives. But then the role of the skilled jazz musician is to creatively build on these basic structures and to improvise new innovations, which help to extend the music that was already there to new horizons. In the same way, thechnology ethics typically tends to build on the foundations provided by traditional and more general ethical therory.»*

Das Problem bei der Anwendung alter Theorien ist, dass diese entwickelt wurden, bevor jegliche moderne Technologien erfunden worden waren, so Nyholm [8]. Ältere Theorien basieren auf der Annahme, dass Menschen mit Menschen interagieren und nicht Menschen mit Technologien. Richtlinien und Gesetze wurden mit der Voraussetzung entwickelt, dass ein Mensch mit einem anderen Menschen interagiert. Ein Beispiel, das er benennt, ist die Frage nach der Verantwortlichkeit eines selbstfahrenden Autos. Die Gesetze des Strassenverkehrs wurden entwickelt, bevor davon ausgegangen wurde, dass ein Auto autonom entscheiden und sogar den Tod eines oder mehrerer Menschen als Folge haben könnte. Nyholm ist der Meinung, dass die veränderten Umstände neue Gesetze und ein neues Denken erfordern.

1.4 KI-Richtlinien und Gesetze

1.4.1 EU AI Act

2021 wurde vom Europäischen Parlament ein Regelwerk zur Regulierung von KI vorgestellt. Der sogenannte EU AI Act soll mittels verschiedener Prinzipien einen einheitlichen Rechtsrahmen zur Entwicklung, Implementierung und Nutzung von KI-Produkten schaffen. KI-Systeme sollen je nach Risikoklassifizierung unterschiedlich stark reguliert werden. Ein System wie ChatGPT ist zum Beispiel einer niedrigeren Risikoklasse zugeordnet als beispielsweise ein selbstfahrendes Auto oder eine medizinische KI. KI-Systeme, die in der EU benutzt werden, sollen laut EU AI Act unter anderem sicher, transparent, nachvollziehbar, nicht diskriminierend und umweltfreundlich sein.[9]

In einem Interview der NZZ sieht Thomas Burri, Professor für Europa- und Völkerrecht der Universität St. Gallen (HSG), in diesem Gesetzestext sehr viel Interpretationsspielraum. Diese Beschlüsse seien von Tech-Expertinnen und Experten festgelegt worden und nicht zivilgesellschaftlich legitimiert.[10]

1.4.2 Situation in der Schweiz

Wie sieht die Regulierung von KI in Bezug auf die Entwicklung und Anwendung in der Schweiz aus? Diese Frage beinhaltet mehrere wesentliche Aspekte: Inwiefern gibt es Gesetze zur Anwendung und Entwicklung von KI im Allgemeinen, wie sind diese formuliert und wie sieht die Situation in Bezug auf die Medizin aus.

In der Blogreihe «Ein nüchterner Blick auf den Mythos KI» der Schweizerischen Akademie der Technischen Wissenschaften sprechen Roger Abächerli (Professor für Medizintechnik) und Thomas Probst (emeritierter Professor für Recht und Technologie, Universität Fribourg) über die Auswirkungen von KI auf den Menschen und die Regulation von KI. Abächerli ist der Meinung, dass Technik unbedingt dem Menschen dienen soll und nicht umgekehrt. Generell sei Angstmacherei jedoch keine gute Idee. Regulierungen und Anforderungen sollen vom Kontext abhängig sein, in dem die KI angewendet wird. Der Staat und Organisationen sollen sich bei Regulierungen nicht unter Druck setzen lassen und dem medialen Hype nicht nachgeben. Auch Probst ist ähnlicher Meinung. Er formulierte einige wichtige Punkte: Vieles ist in der Schweiz bereits rechtlich geregelt. Schnellschüsse gesetzgeberischer Art sollen vermieden werden. Regulation soll dort geschehen, wo wir verstanden haben, was und wo die Probleme der KI sind und es soll keine Überregulation stattfinden. KI soll man nicht idealisieren und nicht vergöttlichen, KI ist dumm und KI versteht nichts.[11]

Die Rolle der FMH Die FMH (Foederatio Medicorum Helveticorum, dt. Verbindung der Schweizer Ärztinnen und Ärzte) hat mittels **Standesordnung** einen Verhaltenskodex für die Schweizer Ärzteschaft geregelt, wie z.B. *Art. 4* der Standesordnung zeigt:

> *Jede medizinische Behandlung hat unter Wahrung der Menschenwürde und Achtung der Persönlichkeit, des Willens und der Rechte der Patienten und Patientinnen zu erfolgen. Arzt und Ärztin dürfen ein sich aus der ärztlichen Tätigkeit ergebendes Abhängigkeitsverhältnis nicht missbrauchen, insbesondere darf das Verhältnis weder emotionell oder sexuell, noch materiell ausgenützt werden. Arzt und Ärztin haben ohne Ansehen der Person alle ihre Patienten und Patientinnen mit gleicher Sorgfalt zu betreuen. Weder die soziale Stellung, die religiöse oder politische Gesinnung, die Rassenzugehörigkeit noch die wirtschaftliche Lage der Patienten und Patientinnen darf dabei eine Rolle spielen.*

Die Standesordnung der FMH führt die moralischen Normen des ärztlichen Handelns an, an die sich jede Ärztin und jeder Arzt der Schweiz zu halten hat. Die in der Standesordnung festgehaltenen Normen sind auch im Bezug auf die Anwendung von KI in der Medizin relevant. Anhand des Beispiels von Art. 4 stellen sich verschiedene Fragen: «[…]Arzt und Ärztin haben ohne Ansehen der Person alle ihre Patienten und Patientinnen mit gleicher Sorgfalt zu betreuen.» Angenommen, eine KI könnte Patienten mit grösserer Sorgfalt, ohne Zeitstress, ohne Vorurteile und mit weniger Fehlern behandeln als ein Arzt oder eine Ärztin. Wäre es dann eine Zuwiderhandlung, wenn sich Ärztinnen und Ärzte nicht der KI bedienen würden?

«Arzt und Ärztin dürfen, ein sich aus der ärztlichen Tätigkeit ergebendes Abhängigkeitsverhältnis, nicht missbrauchen, insbesondere darf das Verhältnis weder emotionell oder sexuell, noch materiell ausgenützt werden.» Wenn eine KI gar nicht erst menschliche Eigenschaften wie Emotionen oder Sexualität besitzen kann, wäre es dann fahrlässig, Menschen auf Patienten loszulassen?

Die beiden Beispiele zeigen, dass sich durch die Entwicklung und Anwendung von KI plötzlich neue Situationen und Fragen ergeben. Sie zeigen, dass neue Richtlinien notwendig sein könnten oder eine Anpassung der bestehenden Richtlinien folgen müsste.

Als Antwort auf die Anwendung von KI in der Medizin hat die FMH die Broschüre «KI im ärztlichen Alltag» erstellt, in welcher sie zehn Forderungen an KI-Systeme stellt, die zu diagnostischen Zwecken herangezogen werden. Mit dem **Forderungskatalog** hat die FMH sich vorgenommen, den durch KI bedingten Wandel in der Medizin mitzugestalten und zu begleiten.

Dieser Katalog fordert unterschiedliche Punkte:

1. KI-Systeme sollen die Zusammenarbeit zwischen Ärztinnen und Ärzten mit Patienten verbessern und nicht ersetzen.
2. KI-Systeme müssen sich an der evidenzbasierten Medizin orientieren.
3. Administrative Prozesse sollen vereinfacht, wirkungsvoll und nutzbringend sein.
4. Es soll eine regelmässige Überprüfung im Sinne einer Post-Market-Surveillance stattfinden.
5. Die Verantwortung bleibt bei den Personen, die die Systeme entwickeln und professionell nutzen; KI-Systeme können momentan nicht verantwortlich gemacht werden.
6. Es ist eine Gebrauchsanleitung erforderlich, damit Ärztinnen und Ärzte die KI-Systeme verstehen, überwachen und gegebenenfalls übersteuern können. Diese soll folgende Punkte beinhalten:
    - Die Rolle der Ärztinnen und Ärzte.
    - Den Zweck bzw. das Krankheitsbild, für welchen/welches die KI eingesetzt werden darf.
    - Die korrekte Verwendung des Systems.
    - Die Methodik, auf der das KI-System beruht.
    - Informationen zur Schulung, Testung und Validierung des Systems.
    - Massnahmen zur Gewährleistung von Datenschutz und Sicherheit.
    - Kontaktstellen für Nutzerinnen und Nutzer bei Problemen.
7. Das medizinische Personal soll kompetent für den Einsatz mit KI ausgebildet werden, und angehende Ärztinnen und Ärzte sollen bereits früh in ihrer Ausbildung damit konfrontiert werden.

8. Die Nutzungsanforderungen sollen sowohl durch Ärztinnen und Ärzte als auch durch Patienten definiert werden.

9. Ärztinnen und Ärzte müssen informiert werden, wenn ihre Arbeit im Hintergrund beeinflusst oder beobachtet wird.

10. KI-Systeme müssen an nationale Datensituationen angepasst werden. Datenschutz und Datensicherheit müssen eingehalten werden, und ein zertifiziertes Qualitätsmanagement der Datennutzung ist erforderlich. Zudem braucht die Schweiz Zugang zu internationalen Datenplattformen.[12]

SAMW Richtlinien Ebenso sollen die Richtlinien der Schweizerischen Akademie der Medizinischen Wissenschaften (SAMW) Ärztinnen und Ärzten medizin-ethische Hilfestellungen bieten. Diese unterliegen einer dauernden Revidierung und es werden ständig neue Richtlinien entwickelt. Zum aktuellen Zeitpunkt (Stand Mai 2025) wurden von der SAMW noch keine Richtlinien zu medizinischen KI-Systemen veröffentlicht.

1.5 Künstliche Intelligenz

1.5.1 Was ist Künstliche Intelligenz?

Obwohl das Konzept der Künstlichen Intelligenz schon seit den 1950er Jahren bekannt ist, ist es schwierig, eine einheitliche Definition für KI zu finden. Die Definitionen sind sowohl in der Wissenschaft als auch in der Gesellschaft sehr unterschiedlich. Künstliche Intelligenz ist einerseits aus dem Wort künstlich, andererseits aus dem Wort Intelligenz zusammengesetzt. Das impliziert, dass etwas eine Art von Intelligenz besitzt, diese jedoch künstlich ist. Darauf folgt die Frage nach der Definition von natürlicher Intelligenz und Intelligenz im Allgemeinen. Die Definition von KI erweist sich unter anderem deshalb als schwierig, da der Begriff Intelligenz weder in der Philosophie noch in der Medizin oder Psychologie einheitlich definiert und verstanden wird.

Mit der Dartmouth-Konferenz 1956 wurde erstmals der Begriff KI erwähnt. Die Konferenz wurde als Workshop mit dem Titel «Dartmouth Summer Research Project on Artificial Intelligence» angelegt. Führende Forscherinnen und Forscher dieser Zeit fanden zur Konferenz in Dartmouth in New Hampshire (USA) zusammen, um den Begriff KI zu erörtern. Das Ganze wurde von McCarthy, Minsky, Rochester und Shannon [13] als «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955» veröffentlicht. Ihre Definition für KI lautete folgendermassen: «KI ist, wenn Maschinen dazu gebracht werden können, jeden Aspekt des Lernens oder jedes andere Merkmal der Intelligenz zu simulieren.» Als Merkmale der Intelligenz nennen McCarthy et al. den Gebrauch der Sprache, die Bildung von Abstraktionen und Konzepten, das Lösen von Problemen sowie Selbstverbesserung. Alles sind Eigenschaften, die dem Menschen bisher vorbehalten waren.

Sven Nyholm, Professor für Ethik der KI, sagte zu KI in seinem Buch «Humans and Robots» folgendes:

> *Sometimes AI is defined as a machine that can behave like a human, sometimes as a machine that can think like a human. […] AI refers to properties of a machine that enable it to perform or imitate tasks a human being would need their intelligence for.*[3]

In der Stanford Encyclopedia of Philosophy wird der Begriff KI folgendermassen beschrieben:

> *Artificial intelligence (AI) is the field devoted to building artificial animals (or at least artificial creatures that – in suitable contexts – appear to be animals) and, for many, artificial persons (or at least artificial creatures) that – in suitable contexts – appear to be persons.*[14]

Die Definition für KI von Spektrum der Wissenschaften lautet:

> *Die künstliche Intelligenz (Abk. KI, engl. artificial intelligence, Abk. AI) ist ein Teilgebiet der Informatik, welches sich mit der Erforschung von Mechanismen des intelligenten menschlichen Verhaltens befasst (Intelligenz). Dieses geschieht durch Simulation mit Hilfe künstlicher Artefakte, gewöhnlich mit Computerprogrammen auf einer Rechenmaschine.*[15]

1.5.2 Gliederung und Arten von KI

KI-Systeme lassen sich einerseits durch ihren Aufbau und die Struktur des zugrunde liegenden Algorithmus klassifizieren, andererseits durch die Intelligenzstufe und die Spannbreite von Aufgaben, die sie zu lösen vermögen. KI-Modelle basieren grundsätzlich auf zwei verschiedenen Ansätzen. Historisch werden symbolische KI und statistische KI unterschieden. Heute basieren die Modelle in der Regel auf statistischen Modellen, welche mittels mathematischen und statistischen Verfahren arbeiten. Hier ist unter anderem das Machine Learning (dt. maschinelles Lernen) angesiedelt, welches das häufigste Modell der KI bildet. Die Begriffe Machine Learning und Deep Learning werden häufig synonym mit dem Begriff Künstliche Intelligenz benutzt. Sie werden häufig allgemein als «Algorithmen» bezeichnet.

Algorithmus Ein Algorithmus bezeichnet die Anleitung für den Computer. Wie das Kochbuch die Anleitung für ein Gericht für eine Köchin oder einen Koch ist, so ist ein Algorithmus die Anleitung für den Computer für ein Programm, das er ausführen soll; er bildet die Anleitung zur Lösung eines Problems. Bei einem «normalen» Programm, das nicht KI-basiert ist, sind alle Strukturen des Algorithmus sehr starr. Sie verändern sich nicht mit der Nutzung. Der Mensch hat immer alle Fäden in der Hand und das Programm wird bei gleichem Input immer den gleichen Output generieren.

Neuronales Netzwerk Viele der KI zugrunde liegenden Konzepte haben ihren Ursprung in der Hirnforschung. Es wurde damit begonnen, algorithmische Strukturen auf Grundlage der funktionellen Struktur des menschlichen Gehirns und somit der Neuronen zu schaffen. Die Netzwerk- und Schichtarchitektur des Gehirns hat die Entwicklung der KI stark geprägt.[16]

Viele KI-Systeme bilden eine aus mehreren «Neuronenschichten» bestehenden Netzwerkarchitektur. In diesen Netzwerken laufen Prozesse, wie das verstärkende Lernen ab - das sogenannte Reinforcement Learning. Systeme lernen durch Versuch und Irrtum und werden für korrekte Lösungen belohnt. Wie das menschliche Gehirn trainiert und lernt ein künstliches neuronales Netzwerk und wird mit der Zeit besser.[17]

Netzwerke wie diese funktionieren so, dass jedes einzelne «Neuron» Signale von der darunter liegenden Schicht erhält und sie an die nächste darüber liegende Schicht weitergibt. In neuronalen Netzwerken können Signale abgeschwächt oder verstärkt werden. Dieser Ablauf führt zu einer Veränderung des Netzwerkes selbst. Ein solcher Prozess wird in der Neurophysiologie als progressive Reifung bezeichnet. Maschinelle neuronale Netzwerke basieren auf linearer Algebra, menschliche Hirnfunktionen hingegen laufen mit biochemischen Prozessen ab.[16]

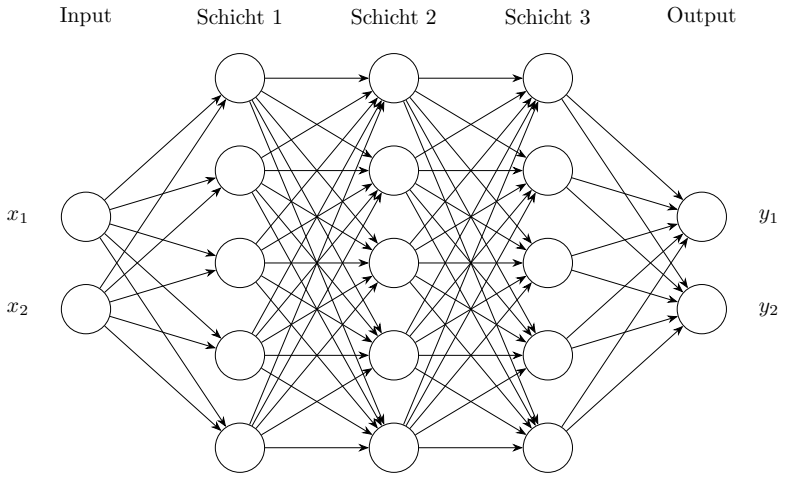

Abbildung 1: Neuronales Netzwerk mit drei Schichten

Machine Learning KI-Systeme funktionieren so, dass sie autonom und ohne menschliche Intervention arbeiten können. Durch Erfahrung und Training können sie lernen, sich mit der Zeit weiterzuentwickeln und werden mit der Zeit allmählich besser. Sie können Ergebnisse und Lösungen generieren, die nicht explizit einprogrammiert wurden. Der Vorgang, aus einer grossen Menge Daten einen neuen Output zu generieren, wird als Machine Learning (dt. maschinelles Lernen) bezeichnet. Machine Learning Algorithmen beruhen auf verschiedenen Ansätzen der Probabilistik (Wahrscheinlichkeitstheorie). Softwares auf Grundlage von Machine Learning arbeiten mit Mustererkennung. Das System sucht in grossen Datensammlungen nach Mustern, Korrelationen und Zusammenhängen und erkennt solche, die für das menschliche Hirn zu komplex gewesen wären.

In «Deep Learning» von Goodfellow, Bengio und Courville [18] beschreibt folgendes Zitat das Vorgehen nach Wahrscheinlichkeiten sehr treffend:

> *In the case of the doctor diagnosing the patient, we use probability to represent a degree of belief, with 1 indicating absolute certainty that the patient has the flu and 0 indicating absolute certainty that the patient does not have the flu.*

Aufgrund des Gelernten können Wahrscheinlichkeitsaussagen getroffen werden. Um gute Ergebnisse zu erhalten, werden aus mathematischen Gründen möglichst grosse Datensätze benötigt.

Deep Learning Deep Learning bildet eine Untergruppe des Machine Learnings. Deep Learning ist eine stärkere Form der KI, die sich vor allem dadurch unterscheidet, dass sie ein komplexeres und vielschichtigeres neuronales Netzwerk besitzt als Machine Learning Algorithmen. Sie vermögen grössere Datensätze zu analysieren. Diese Systeme finden beispielsweise in der Bild- und Spracherkennung ihren Nutzen, bei denen es sich um riesige Datenmengen handelt.[19]

Überwachtes und unüberwachtes Lernen Bei Machine Learning und Deep Learning Systeme, die mit Daten trainiert werden, die gelabelt sind, wird vom sogenannten überwachten Lernen gesprochen. Das bedeutet, dass die Daten mit Informationen etikettiert sind, also ein Label besitzen. Wenn das System selbstlernend ist und es Muster und Strukturen selbständig in nicht-gelabelten Daten erkennt, wird dieser Vorgang als unüberwachtes Lernen (unsupervised learning) bezeichnet.[16]

Intelligenzstufen von KI Diese Unterscheidung basiert auf der Ausprägung der Intelligenz eines Systems. Verschiedene KI-Systeme lassen sich dabei auf einem Spektrum abbilden. Ein Ende des Spektrums bilden KIs, die nur eine einzige Aufgabe lösen können. In diesem Fall wird von einer schwachen KI (engl. weak AI) gesprochen. Ein Beispiel dafür ist der Schachcomputer Deep Blue. Deep Blue schlug 1996 den amtierenden Schachweltmeister Garri Kasparow in einem Schachspiel. Eine Schach-KI beherrscht einzig und allein das Schachspielen gut, diese eine Aufgabe sehr gut, besser als je ein Mensch dies könnte. Eine schwache KI ist in einer spezifischen Aufgabe enorm ausgeprägt.[20] Das andere Ende des Spektrums bilden KIs, die jede denkbare Aufgabe zu lösen vermögen. Sie werden als allgemeine künstliche Intelligenz (engl. Artificial General Intelligence, AGI) oder als starke KI (engl. strong AI) bezeichnet. AGI ist das, was wir aus vielen Science-Fiction-Filmen kennen. Eine AGI wäre folglich dem Menschen in allem überlegen. Von einer AGI, wie wir sie aus dem Kino kennen, sind wir aktuell jedoch noch weit entfernt. Sie existiert momentan nur theoretisch.[8]

1.5.3 KI in der Medizin

In der Medizin fallen täglich unzählige Daten an. Jeder Patient wird in Datenpunkte erfasst wie beispielsweise die ärztliche Krankengeschichte, Laborwerte, Bilddaten oder digitale Herzrhythmusüberwachungen einer Smartwatch. Je mehr die Digitalisierung im Gesundheitswesen fortschreitet, desto mehr Daten fallen an. Diese Daten helfen einerseits, mehr über einen Patienten zu erfahren, diesen besser kennenzulernen und möglicherweise bessere Diagnosen zu stellen. Andererseits stellen sie uns aber auch vor die Herausforderung, riesige Datenmengen adäquat verarbeiten und auswerten zu können. Fehlt die Übersicht über die Daten und die Möglichkeit, diese zu nutzen, versinken wir in einer Datenflut, die weder den Ärztinnen und Ärzten noch dem Patienten weiterhelfen.

Der Arzt William Schwartz war in seinem Artikel im New England Journal of Medicine bereits vor 50 Jahren der Meinung, dass in Zukunft Ärztinnen und Ärzte und Computer häufig miteinander sprechen und kommunizieren würden. Der Computer würde Notizen zum Verlauf, zu Labordaten oder Ergebnissen automatisch anlegen und den Ärztinnen und Ärzten mit Hinweisen zu Diagnosen behilflich sein. Fünfzig Jahre später sind wir kaum weiter. In vielen Gebieten in der Wirtschaft oder in der Industrie ist die Digitalisierung bereits viel weiter fortgeschritten. Die Medizin steckt noch immer in der Anfangsphase fest. Nach wie vor wird in vielen Praxen oder Spitälern vieles auf Papier festgehalten. Patientenmanagementsysteme sind von Spital zu Spital unterschiedlich und der Datenaustausch ist mühsam. Sowohl Patientendossiers als auch Bilddateien sind häufig inkompatibel. Aus der Perspektive von Schwartz scheint es unwahrscheinlich, dass KI das Gesundheitswesen in naher Zukunft im Grundsatz revolutionieren wird.[21]

Entgegen dieser Aussage gibt es immer mehr Software und Tools, die für die Medizin entwickelt werden. Insbesondere in der bildbasierten Diagnostik, beispielsweise in der Radiologie, Pathologie, Ophthalmologie oder Dermatologie, gibt es bereits viele Ansätze von KI-Systemen. In der Radiologie kommt KI zum Zuge, wenn es um die Auswertung von CT-, MRI- oder Röntgenbildern geht. KI-Algorithmen sind auf Mustererkennung spezialisiert, was der Grund dafür ist, dass diese bis jetzt vor allem in der Bilderkennung genutzt werden. Aufgrund der Entwicklung von Large Language Models[2] ist auch die Erkennung von natürlicher Sprache und Text mittlerweile einfacher geworden. Dadurch werden medizinische Chatbots anwendbar und Patientenakten interpretierbar gemacht, und es ermöglicht uns auch, Daten in Text umzuwandeln. Ebenso hilfreich erweist sich KI in der personalisierten Medizin. KI kann helfen, anhand von molekularen, genetischen, klinischen und laborchemischen Daten eine massgeschneiderte Behandlung für Patienten zu liefern, anstatt eines Medikaments, das in der gleichen Dosierung für alle Patienten mit der gleichen Erkrankung verwendet wird. Somit kann der Behandlungserfolg gesteigert und Nebenwirkungen vermindert

[2]Large Language Models (LLMs) sind KI-Systeme, die auf natürlicher Sprache basieren. Diese basieren auf neuronalen Netzwerken und werden mit grossen Datenmenen an natürlicher Sprache trainiert. In der Regel werden dafür Texte, wie z.B. Wikipediaartikel, verwendet. Ein Beispiel für ein LLM ist Chat-GPT.

werden. Auf molekularer Basis können neue Wirkstoffe entwickelt werden, was die Entwicklung von neuen Medikamenten einfacher und günstiger macht.

Catherine Jutzeler, Assistenzprofessorin Departement Gesundheitswissenschaften und Technologie der ETH ist im Gegensatz zu William Schwarz der Meinung, dass KI in der Medizin in den nächsten Jahren eine grosse Rolle spielen wird. Die Richtung der Anwendung von KI in der Medizin wird sich jedoch erst noch zeigen, je nachdem, wie die Gesellschaft und die Politik diese Entwicklung gestalten werden.[22]

1.5.4 Ethische Fragestellungen in der Anwendung von KI in der Medizin

Aufgrund der vielen Vorteile und der breiten Anwendungsmöglichkeiten schreitet die Implementierung von KI-Systemen im Gesundheitswesen rasant voran. Bei genauerem Hinsehen eröffnet sich jedoch ein grosses Feld an ethischen und rechtlichen Fragestellungen.

Aufgaben und Entscheidungen werden durch die Anwendung von KI zunehmend an Computer abgegeben. Dadurch wird auch Verantwortung übertragen. Wenn ein System einwandfrei funktioniert und gute Ergebnisse liefert, ist das kein Problem. Die Frage nach Verantwortung stellt sich erst, wenn ein Produkt nicht wunschgemäss funktioniert. In der Medizin bedeutet eine Fehlfunktion im schlimmsten Fall eine Fehldiagnose oder eine verpasste Diagnose, eine falsche Behandlung oder ein falsches Medikament. Da die KI ein Teil im Entscheidungspfad geworden ist, verändert sich die Zuweisung der Verantwortung und Haftung. Kann eine KI Verantwortung übernehmen oder soll diese sowohl bei der Entwicklerin und beim Entwickler als auch beim Anwender liegen? Durch die Nutzung von KI-Systemen entstehen zusätzlich neue Verantwortungslücken, Situationen, in denen unklar ist, wer oder was die Verantwortung trägt. Gewisse Verantwortungslücken waren bei genauer Betrachtung im Gesundheitswesen schon immer vorhanden, werden durch KI jedoch zusätzlich verstärkt.

Eine weitere Überlegung, die aus der Anwendung von KI resultiert, ist jene nach einer ethischen Verpflichtung der Nutzung dieser Systeme. «Arzt und Ärztin benützen die ihnen angebotenen Möglichkeiten zur Sicherung der Qualität ihrer Arbeit. Sie sind zur ständigen Fortbildung gemäss Fortbildungsordnung verpflichtet.»[23] An diesen Grundsatz der Standesordnung der FMH, hat sich jeder Arzt und jede Ärztin zu halten. Liefert eine KI bessere Resultate, Diagnosen oder Behandlungsvorschläge, operiert eine KI besser als ein Arzt oder eine Ärztin, könnte ein Verzicht auf die Nutzung von KI eine Verletzung dieses Grundsatzes darstellen.

KI-Systeme funktionieren über eine Unmenge an Daten. Medizinische Daten zu sammeln bedeutet, Daten und Informationen von Menschen zu erfassen. Der Prozess der Sammlung und Verarbeitung sowie der daraus resultierenden Anwendung der datenbasierten Programme birgt Gefahren. Der sichere Umgang mit Patientendaten sowie der bewusste Umgang mit denselben sind wichtig, um die Gerechtigkeit und Privatsphäre der Patienten zu wahren. Je nachdem, wo und wie Daten gesammelt werden, entstehen meistens Verzerrungen, sogenannte Bias. Wenn zum Beispiel CT-Bilddaten in einem Spital in der Schweiz gesammelt werden und daraus ein KI-System

programmiert wird, wird das System in einem anderen Land oder einem anderen Spital nicht direkt anwendbar sein. Das ist vor allem dann ein Problem, wenn eine KI auf unterrepräsentierte Patientengruppen angewendet wird.

Zu verstehen, wie eine KI einen Output generiert und wie dieser zustande kommt, ist komplex. Ein Deep Learning Algorithmus lernt selbständig aus den eingegebenen Daten und errechnet so einen eigenen Lösungsansatz. In der Praxis ist nicht mehr jeder Schritt erklärbar. Dieses Phänomen wird als Black Box bezeichnet. Es wird ein Input eingegeben und ein Ergebnis ausgegeben. Wie dieses entstanden ist, ist nicht mehr komplett transparent. Würde der komplette Vorgang eingesehen werden können und wäre er vollständig nachvollziehbar, würde von einer «glass box» gesprochen werden. Müssen Ärztinnen, Ärzte und Patienten komplett nachvollziehen und verstehen können, wie ein Ergebnis entsteht? In der Regel verstehen eine Ärztin oder ein Arzt auch ohne KI nicht im Detail, wie ein Computer oder ein Laboranalysegerät funktioniert, und doch wird täglich damit gearbeitet und darauf vertraut.

Immer wichtiger wird auch das Thema Ressourcen. Einerseits steigt der Energieverbrauch der Menschen auf diesem Planeten jeden Tag und doch wollen wir gleichzeitig eine nachhaltigere Lebensweise schaffen. Die KI hat das Potenzial, Ressourcen zu sparen. Dennoch darf nicht vergessen werden, dass die Menge an verarbeiteten Daten exponentiell ansteigt und damit auch der Energieverbrauch. Die jährlich steigenden Krankenkassenprämien werfen zudem die Frage nach den finanziellen Ressourcen auf. Das Gesundheitswesen muss in den nächsten Jahren oder Jahrzehnten von Grund auf reformiert werden. KI könnte zur Lösung dieses Problems beitragen. Der Einsatz von KI könnte jedoch ebenso das Gegenteil bewirken. Die Gesundheitskosten könnten ansteigen, ohne dass eine Verbesserung des Gesundheitssystems eintritt.

Wenn man vor der Wahl steht, sich von einem Operationsroboter operieren zu lassen oder von einem Chirurgen oder einer Chirurgin, während der Roboter eine niedrigere Fehlerquote hat als ein (menschlicher) Chirurg, würde man dann lieber vom Roboter oder vom Menschen operiert werden? Welche Relevanz hat die Zwischenmenschlichkeit und Empathie einer Ärztin oder eines Arztes am Behandlungserfolg und der Patientenzufriedenheit? In welche Richtung wird sich der Arztberuf entwickeln?

Diese Arbeit konfrontiert Expertinnen und Experten aus dem Bereich der medizinischen KI-Entwicklung mit diesen Fragen und veranschaulicht deren individuelle Überlegungen zu den moralischen und ethischen Aspekten des Themas.

2 Methodik

In diesem Abschnitt wird das methodische Vorgehen dieser Arbeit beschrieben. Die Methodik wurde entsprechend der Forschungsfrage ausgewählt. Für diese Studie wurde qualitativ vorgegangen und es wurden semistrukturierte Interviews geführt. Für die Auswertung wurden die Interviews entweder mittels Online-Plattform «Zoom» oder der iPhone-App «Voice Recorder» aufgezeichnet und mittels der Transkriptionssoftware «Happy Scribe» transkribiert. Die Transkripte wurden anhand des interpretativen phänomenologischen Analyseansatzes (IPA) ausgewertet. Bei dieser Arbeit wurde induktiv vorgegangen.

2.1 Forschungsdesign

2.1.1 Qualitative Methodologie

Nach Flick, Kardorff und Steinke [24] bedeutet qualitative Methodologie, dass Handlungen und Denkweisen aus Sicht der befragten Menschen betrachtet werden. Ziel ist, dass die Ansichten der Befragten kontextualisiert werden. Aussagen müssen zuerst in den sozialen und kulturellen Kontext der befragten Person gestellt werden, um so auch grössere Zusammenhänge zu verstehen. Es wird versucht, ein Verständnis für die Aussagen und Ansichten der Personen zu erarbeiten und deren komplexe Zusammenhänge zu verstehen, anstelle diese quantitativ zu verarbeiten. Fragen werden in der qualitativen Forschung offen formuliert, damit den Perspektiven der Personen viel Raum gelassen werden kann und sie nur wenig durch die Fragen eingeschränkt werden.

Die qualitative Forschungsmethodik wurde ausgewählt, da sie sich dazu eignet, die Lebenswelten von Personen (Interviewpartnern) «von innen heraus», das bedeutet aus der Sicht des handelnden Menschen, zu beschreiben. Qualitative Forschung ist im Gegensatz zu quantitativen Methoden näher an den untersuchten Phänomenen von Individuen.[24] Im Bezug auf diese Arbeit ist dieser Punkt relevant, da die individuellen Ansichten der Entwicklerinnen und Entwickler zu KI in der Medizin und deren ethischen Aspekten relevant sind. Um die persönlichen Meinungen bestmöglich zu erörtern, wurde auf die qualitative Herangehensweise zurückgegriffen.

Als methodologische Rahmenstruktur wird mittels phänomenologischer[3] Lebensweltanalyse gearbeitet. Dieser Ansatz und weitere Aspekte der qualitativen Forschung werden im Folgenden genauer erläutert.

[3]Phänomenologie ist ein zentrales Konzept in der qualitativen Forschung. Sie ist sowohl eine Forschungsmethode, als auch ein philosophischer Ansatz. In der Forschung ist das Ziel der Phänomenologie Erfahrungen so zu erkennen und zu beschreiben, wie sie von einem Individuum erlebt werden. Erfahrungen und das Erleben der Welt sind gemäss der Phänomenologie subjektiv. Die Phänomenologie versucht eine möglichst feie und unvoreingenommene Erfassung der Wirklichkeit der betrachteten Person zu erreichen, auch wenn das nie komplett möglich ist. Das Ziel ist es, die Bedeutung des Gesagten, aus Sicht und im Kontext der Umgebung der betrachteten Person zu verstehen.[5]

2.1.2 Interpretativer phänomenologischer Analyseansatz

Die interpretative phänomenologische Analyse (IPA) versucht, das subjektive Erleben von Individuen zu erkennen und zu verstehen. Das bedeutet genauer, welchen Sinn die Individuen in ihrem Erlebten sehen. Die IPA stützt sich auf die fundamentalen Prinzipien der Phänomenologie, Hermeneutik[4] und Idiographie[5]. Vereinfacht gesagt, fokussiert sich die IPA darauf, in den Schuhen des untersuchten Subjekts zu stehen und zu verstehen, was das bedeutet. Dies geschieht als dynamischer Prozess, indem die Forscherinnen und Forscher durch interpretative Aktivität eine aktive Rolle einnehmen. Es handelt sich um eine doppelte Hermeneutik. Der Untersuchte macht Sinn aus seiner erlebten Welt und der Untersucher versucht, diesen Sinn zu verstehen, die Welt des Untersuchten aus dessen Augen zu sehen und nicht nur zu beschreiben. Die IPA beschäftigt sich eher mit dem Einzelnen als mit dem Universellen.[25] Dieser Ansatz eignet sich daher gut zur Beantwortung der Forschungsfrage, da subjektive Meinungen und das innere Erleben der Untersuchten eine zentrale Rolle in dieser Arbeit einnehmen.

2.2 Empirische Ethik

Um die moralischen und ethischen Überlegungen von Individuen veranschaulichen zu können, findet die empirische Ethik oft Anwendung in der Medizinethik. Sie versucht, mittels empirischer Untersuchung anhand sozialwissenschaftlicher Methoden (in dieser Arbeit anhand Interviews) Daten zu erheben. Diese werden dann mittels normativer ethischer Diskussionen verarbeitet.[26] In dieser Forschungsarbeit wird mittels empirischer Ethik untersucht, wie Expertinnen und Experten aus dem Bereich der KI-Entwicklung die Implikationen und Entwicklungen von KI in der Medizin wahrnehmen. So können anhand dieses Ansatzes ethische Grundsätze entdeckt und weiterverfolgt werden.

2.3 Auswahl der Interviewteilnehmerinnen und Interviewteilnehmer

Für eine umfassende Beantwortung der Forschungsfrage wurden Teilnehmerinnen und Teilnehmer gesucht, die fundierte und relevante Erfahrungen in Data Science, Software Engineering, Digital Medicine, Computer Science, Biomedical Engineering oder Data Analytics verfügen und an der Entwicklung von medizinischen Technologien mit KI beteiligt sind. Die Auswahl erfolgte nach dem Prinzip des zielgerichteten Samplings, bei dem gezielt einzelne Personen ausgewählt werden. Die ausgewählten Personen wurden anschliessend durch E-Mails rekrutiert.

[4] Hermeneutik (aus dem Altgriechischen «erklären, auslegen, übersetzen»), meint die Kunst des Verstehens, Interpretierens und Sinngebens eines Textes. In der Hermeneutik versucht man einen Text zu verstehen und diesem Sinn zu geben.[5]

[5] Der idiographische Ansatz, ist eine erkenntnistheoretische Methode, bei dem die betrachteten Personen möglichst individuell und in ihrem Kontext angeschaut werden. Wichtiger als ein generelles Statement, wie sonst in der empirischen Forschung, ist eine genaue Einzelfallanalyse.[25]

In der IPA ist die Anzahl an Teilnehmerinnen und Teilnehmern im Vergleich zu anderen Methodologien tendenziell klein. Für die vorliegende Fragestellung ist eine grosse Stichprobenanzahl unpassend, da eine tiefgründige Untersuchung von Individuen vorgenommen wird. Um diesen Interviews gerecht zu werden, wird von einer empfohlenen Teilnehmeranzahl von sechs bis acht Probanden ausgegangen.[25]

Um ein möglichst diverses und fundiertes Bild zu erhalten, wurden sieben Teilnehmerinnen und Teilnehmer interviewt. Durch die verschiedenen Perspektiven konnte eine thematische Saturation erreicht werden.

2.4 Datenerhebung

2.4.1 Semistrukturierte Interviews

Um detaillierte, reichhaltige und persönliche Erfahrungsberichte zu erheben und direkte Einblicke in die subjektiven Erfahrungen und Phänomene der Probanden zu erhalten, sind semistrukturierte Eins-zu-eins-Interviews die Methode der Wahl. So kann in Echtzeit ein Dialog stattfinden, der sowohl Raum und Flexibilität als auch Struktur erlaubt.[25]

Ziel der Arbeit ist es, die subjektiven und ethischen Überlegungen, die im Zusammenhang mit der Entwicklung von KI-Technologien in der Medizin eine Rolle spielen, zu erfassen und das entsprechende Expertenwissen in Erfahrung zu bringen. Kommunikation und Dialog sind bedeutende Elemente dieser Methode und elementar in der qualitativen Forschung. Durch die semistrukturierten Interviews wird die Erforschung von einzelnen Sichtweisen und dadurch das Generieren von neuen Hypothesen möglich.[26] Damit konnten sowohl offene Fragen gestellt werden, um eine möglichst breite Varianz an Antworten zu erhalten, als auch auf die gegebenen Antworten eingegangen und Folgefragen gestellt werden. Auf diese Weise konnte die Phänomenologie des Einzelnen berücksichtigt werden. Gleichzeitig blieb der Kern der Arbeit erhalten und die Ursprungsfrage wurde beibehalten.

Ausgehend von der Fragestellung wurde ein Interviewleitfaden entwickelt. Daraus entstanden fünf Teilbereiche. Der Interviewleitfaden deckte grob folgende Gebiete ab:

1. Einstieg und persönlicher Hintergrund
2. Persönlicher Bezug zu KI
3. Persönliches Verständnis von Ethik und Ethik im Bereich von KI
4. Ethik im Kontext KI in der Medizin
5. Zukunft von KI-Anwendung in der Medizin

Um auf die Antworten der Probanden genauer eingehen zu können, lag zusätzlich ein Nachfrageteil vor. Dieser ermöglichte es, je nach Antwort unterschiedliche Themengebiete vertiefend zu behandeln. Es wurden zwei Probeinterviews durchgeführt, um die

Verständlichkeit, Sinnhaftigkeit sowie die Interviewdauer zu testen. Die Probeinterviews flossen nicht in diese Analyse mit ein. Die Grundversion des Interviewleitfadens blieb bis zum Schluss erhalten, jedoch wurden die Fragen während des Interviews je nach Proband leicht abgeändert oder den vorangegangenen Antworten angepasst. Die Grundstruktur des Interviewleitfadens befindet sich im Anhang. Sechs Interviews fanden über die Videokonferenzplattform «Zoom» statt und wurden direkt darüber aufgezeichnet. Ein weiteres Interview erfolgte vor Ort und liess sich mittels App «Voice Recorder» auf dem iPhone aufnehmen. Die Interviews dauerten jeweils zwischen 50 und 65 Minuten. Vorgängig wurde von allen Probanden eine Einverständniserklärung eingeholt, die das Aufzeichnen des Interviews, den Gebrauch der Daten sowie die Anonymisierung behandelte. Alle Interviews wurden verschlüsselt, indem die Namen anonymisiert und durch die Ziffern I–VII ersetzt wurden.

2.5 Transkription

Um das gesprochene Interview für die wissenschaftliche Analyse dauerhaft haltbar zu machen und zu analysieren, ist die Transkription notwendig.[24] Für die Transkription wurde die Transkriptionssoftware «Happy Scribe» verwendet. Im Anschluss folgte eine manuelle Überprüfung und bei Bedarf eine Korrektur, um eine möglichst hohe Authentizität und Genauigkeit zu gewährleisten. So konnte eine präzise Wiedergabe der Äusserungen sichergestellt werden. Verwendet wurde die Standardorthographie der entsprechenden Sprache. Zwei Interviews fanden auf Schweizerdeutsch, vier auf Englisch und eines auf Hochdeutsch statt. Die originalen Äusserungen und Sprachen bleiben erhalten, Schweizerdeutsch wurde zwecks Anonymisierung und besserer Lesbarkeit ins Hochdeutsche übertragen.

2.6 Analyse der Daten

Die Analyse der Leitfadeninterviews geschah mittels IPA, die teilweise schon in *Abschnitt* 2.1.2 beschrieben ist. Das Konzept der IPA kann im konkreten Fall sehr unterschiedlich umgesetzt werden. Grundsätzlich wird in mehreren Schritten gearbeitet. Die Analyse erfolgte unter anderem nach Pietkiewicz und Smith [25] und anhand des Kapitels «Analyse von Leitfadeninterviews» in Flick, Kardorff und Steinke [24] von Christiane Schmidt.

Die folgenden Schritte zeigen das Vorgehen der IPA. Sie sind nicht starr, sondern bilden einen flexiblen Rahmen. Das bedeutet, dass man zwischen den Schritten hin- und herwechseln kann, um die Analyse zu verfeinern und zu verbessern.

1. Mehrfaches lesen und Notizen machen

2. Erste Codes generieren: Es wird systematisch nach Schlagwörtern für prägnante Textstellen gesucht. Kurze Begriffe, Wörter oder Phrasen sollen diese Textstelle bezeichnen und markieren - die sogenannten Codes.

3. Entwicklung von Themen: Aus den Codes werden passende Gruppierungen entwickelt, damit die Textpassagen thematisch geordnet werden können, dabei entstehen initial Unterkagegorien und daraus Kategorien.

4. Die herausgearbeiteten Kategorien werden mittels Zitaten aus den Interviews belegt.

5. Themen und Muster werden über verschiedene Interviews verglichen.

In Punkt 1 werden Auswertungskategorien erstellt, indem Begriffe und Aspekte zu Textpassagen notiert werden, die im weitesten Sinne auf die Fragestellung eingehen. Wichtig ist eine möglichst offene Herangehensweise, um zu erkennen, ob die Befragten die Begriffe der Fragestellung aufgreifen und welche Bedeutung diese für sie haben. Durch den Prozess des Codierens entstehen Kategorien und Unterkategorien, wodurch sich der Codierleitfaden ergibt. Der Codierleitfaden ist eine systematische Übersicht der extrahierten Kategorien und Unterkategorien. Den Kategorien und Unterkategorien können anschliessend entsprechende Textpassagen passend zugeordnet werden. Dieser Prozess ist beispielhaft in *Tabelle* 2 dargestellt.[25]

Zitat	*…explainability is not a requirement in my perspective that is universal. So something is always a black box, depending on the person approaching the box […] a black box is not something bad per se and comes in degrees, in my opinion.*
Initialer Code	Transparenz, bzw. Black Box kommt in Gradierungen, ist nicht generell etwas Schlechtes
Unterkategorie	Nachvollziehbarkeit
Kategorie	Verantwortungsvolle Entwicklung und Implementierung von KI-Systemen

Tabelle 2: Beispielhafter Codierungsprozess

Das Ziel dieses Ansatzes ist der Austausch zwischen Material und theoretischem Vorverständnis. Daraus werden dann Auswertungskategorien und schlussendlich eine vertiefte Fallinterpretation erstellt. Allenfalls ergibt sich dann eine neue Hypothese oder eine neue theoretische Überlegung.

2.7 Induktion

Diese Arbeit wurde induktiv angegangen. Die Interviews wurden ohne vorhergehende Theorien oder Hypothesen analysiert. Aus diesen Daten wurden Themen und Kategorien erstellt, woraus Theorien und Schlüsse entwickelt werden können.

2.8 Strategische Literatursuche

Eine umfassende systematische Literatursuche wurde durchgeführt, um die Einleitung und die Diskussion zu unterstützen. Für die Entwicklung des Interviewleitfadens diente aktuelle Literatur als Grundlage. Die Recherche erfolgte gemäss den PRISMA Guidelines.[27] Die elektronische Suche fand in den Online-Datenbanken PubMed, Medline (Ovid) und Cochrane Library statt. Ergänzend dazu erfolgte eine manuelle Suche nach zentralen Artikeln und Übersichtsarbeiten, um weitere relevante Quellen zu identifizieren. Die Suche erfolgte unter Verwendung der folgenden Schlüsselwörter:

```
(((AI[Title]) OR (Artificial Intelligence[Title]))
AND (ethics[Title])) AND ((healthcare) OR (medicine.))
```

3 Ergebnisse

In diesem Abschnitt erfolgt die Auswertung der geführten Interviews. Durch die Analyse nach IPA entstanden aus den Interviews vier Hauptkategorien. Wie in *Abschnitt* 2.1.2 Interpretativer phänomenologischer Analyseansatz und in *Abschnitt* 2.6 Analyse der Daten beschrieben, wurden diese Hauptkategorien aus den definierten Codes erstellt. Weiter ergaben sich mehrere Unterkategorien. Im folgenden Abschnitt werden die Hauptkategorien und Unterkategorien erläutert und mit Zitaten der Interviews belegt, welche in hochdeutscher und englischer Sprache verschriftlicht wurden.

3.1 Verantwortungsvolle Entwicklung und Implementierung von KI-Systemen

Der erste Abschnitt behandelt die verantwortungsvolle Entwicklung und Implementierung von KI-Systemen.

3.1.1 Ehrliche und Transparente Kommunikation

Es zeigte sich in mehreren Interviews, dass es aus Sicht der Entwicklerinnen und Entwickler wichtig ist, offen nach aussen zu kommunizieren. Dies kann unterschiedliche Kommunikationsformen und Wege umfassen, wie etwa den Austausch mit der Gesellschaft oder der Forschung. Zudem gehört zur guten Kommunikation auch, Daten ehrlich und offen zu präsentieren. In diesem Kontext wurde eine transparente Kommunikation als besonders wichtig erachtet:

> *For instance, one has to be very honest about the data source and the biases in the data. And then, what are the ethical aspects of it?*

Kommunikation bedeutet auch, ganz klare und genaue Angaben über die entwickelten Programme geben zu können. Damit ist gemeint, klar zu definieren, was ein Algorithmus kann und was nicht.

> *Let's imagine we are in Bern and you sample data from people in the main station in Bern. […] But what does it mean in terms of the applicability of your AI? Probably you should refrain from using your AI in Polynesia or in Indonesia because it's simply not done in that context.*

Diese Ehrlichkeit gegenüber der Performance des entwickelten Algorithmus zeigt sich auch in folgendem Zitat sehr gut:

> *Man muss auch sagen, viele Publikationen sind mit Vorsicht zu geniessen. Wenn zum Beispiel gesagt wird, ein Modell performe auf diesen Daten sehr gut, dann muss man fragen: «Okay, aber wie sind die Daten?»*

Bezug auf das Präsentieren von beschönigten Daten beziehungsweise auf die nicht ganz transparente Offenlegung der eigenen Resultate nimmt der Begriff «cherry-picking», der mehrfach erwähnt wurde. Cherry-picking meint das Aussuchen von einzelnen, erwünschten, guten Daten und das Auslassen der schlechten oder erfolglosen Daten.

> *Or when you publish a paper with results, you also have to follow the ethics of you don't want to cheat on the results, you don't want to cherry-pick, you want to be fair with the community, you don't want to create a fake inflated set of results that then no one can arrive at.*

Aus diesen Aussagen folgt: Entwicklerinnen und Entwickler machen sich Gedanken zur Kommunikation. Eine ehrliche und sehr offene Darlegung der Resultate ihrer eigenen Algorithmen ist erforderlich.

3.1.2 Umgang mit Systemfehlern

Ein sehr häufig genannter ethischer Aspekt ist der Umgang mit Systemfehlern. Was passiert, wenn die eigenen Systeme versagen oder zu Fehldiagnosen führen? Das ist ein Thema, das mehrfach Erwähnung fand und eine Frage, die offenbar bei vielen präsent ist und Bedenken auslöst, wie die folgenden Zitate zeigen.

> *For the moment, the tools that I see, at least in pathology, have a lot of improvement that is absolutely necessary.*

Den meisten ist bewusst, dass Fehler passieren können. Diese können in unterschiedlichem Ausmass geschehen.

> *Gerade wenn wir mit Black-Box-Modellen arbeiten, können die Failure-Modes so katastrophal und unerwartet sein.*

Die Vorstellung eines Systemversagens eines eigenen oder mitentwickelten Algorithmus scheint bereits zu unangenehmen Gefühlen zu führen.

> *If the doctor would have relied directly on it, and I know that there was a mistake, I would feel devastated, completely devastated to have that responsibility on myself. On this, it would be really, really difficult to live with that thing, to know that the thing you did caused harm to somebody.*

Systemfehler werden auch als Albtraumsituationen beschrieben.

> *That's a nightmare situation (system failure) because you know there's this hope, again, going back to the data-driven process where you hope that there's something in that data that connects your whatever input and the output of the model. It's a leap of faith. You hope that that thing exists, that correlation.*

Fehler sind ein integraler Teil der Entwicklung sowie der Implementierung von Algorithmen. Obwohl Fehler auftreten, wurde wiederholt betont, dass das Verständnis dieser Fehler und die gezielte Vermeidung ihrer Wiederholung von entscheidender Bedeutung sind, wie die folgenden Zitate zeigen.

> *Also es wird natürlich immer Fehler geben. Das ist auch etwas, was man akzeptieren muss.*

> *In the case for the person who's developing it, you want to understand what went wrong for this patient. So that patient's data is crucial to be able to understand the defect and the algorithm so that we can include it and not let that happen again and increase the data that are similar to that patient's data so this doesn't happen again.*

Als Problematik wird erwähnt, dass Entwicklerinnen und Entwickler weit weg von der Anwendung am Patienten sind.

> *Emotional bin ich so weit weg vom Patienten, dass ich nie in dem Sinne positive oder negative Rückmeldungen erhalte.*

3.1.3 Transparenz und Nachvollziehbarkeit

Das Wort «Black Box» erwies sich als allgegenwärtig. Es scheint ein Begriff zu sein, welcher nicht von allen gleich, jedoch von den meisten ähnlich verstanden wird.

> *Es ist ein wenig wie ein Heuhaufen: Auch wenn du weisst, nach was du suchst, darin zu suchen, geht sehr lange. In diesem Sinne ist alles transparent. Du kannst eigentlich alles ansehen, aber wirklich nachvollziehen, wieso was gemacht wurde, das ist schwierig. Black Box tönt so ein wenig wie eine Schachtel, in die du nicht hineinschauen kannst. Aber du kannst hineingucken, das Problem ist nur, die Komplexität ist so gross.*

> *Yes, everything is a black box, because it depends on the person who approaches the object.*

> *Die Problematik ist, dass möglicherweise höhere Dimensionen von Strukturen im Spiel sind, aber wir Menschen denken 3D, vielleicht 4D, mit der Zeit.*

«Shortcut Learning»[6] ist ein Begriff, der mehrmals gefallen ist. Dies scheint ein wichtiger Faktor zu sein, der während der Entwicklung beachtet werden muss.

> *There is a phenomenon that's super important and interesting. It's called shortcut learning. […] Shortcut learning is basically your AI system.[…] The problem is that this type of strategy most likely will not generalize. So when I move my system to another device, to another hospital, it might fail because that other system might not have that little marker, and so on.*

> *…dass in unseren Daten, die gebraucht worden sind, irgendwelche Artefakte drin sind, die zulassen, dass das Modell Abkürzungen findet und gar nicht eigentlich auf das zu schauen, auf was man will.*

Die Nachvollziehbarkeit scheint nicht für alle Beteiligten gleich möglich und nicht gleichermassen nötig zu sein.

> *So explainability is not a requirement in my perspective that is universal.*

Hier wird häufig auf die Rolle der Ärztin und des Arztes eingegangen und inwiefern der Algorithmus für diese transparent sein soll.

[6] Shortcut Learning bezeichnet ein Phänomen, bei dem ein KI-Modell eine Abkürzung zur Lösung einer Aufgabe findet. Der abgekürzte Lösungsweg führt oft zu richtigen Ergebnissen, ist aber nicht generalisiert und basiert auf falschen Annahmen. Das folgende bekannte Beispiel illustriert das Problem sehr gut: Eine KI-Software, hat das Ziel, auf Röntgenbildern, den Schweregrad einer Erkrankung zu erkennen, anstelle von physiologischen und pathologischen Begründungen, findet die KI eine sogenannte Abkürzung. Sie hat gelernt, dass Patienten mit vielen Kabeln, Kathetern und Schläuchen kranker sind und gesunde Patienten in der Regel weniger Zugänge und Kabel haben. Das heisst je mehr Kabel, desto schwerer die Krankheit. Warum ist das problematisch? Die KI bewertet nicht die Erkrankung des Menschen, sondern zählt Schläuche.

AI systems should be able to have a conversation mode, where you have this bot that's telling you, look, I analyze the data and this seems to be the better way to treat this patient. But you, as a doctor, say, well, show me some facts on this because I disagree with you, which happens in medicine, right?

Die Befragten erwarteten in der Regel ein gewisses Verständnis der Ärztinnen und Ärzte in Bezug auf die Voraussagen des Systems. Dies setzte eine gewisse Transparenz voraus, jedoch kein komplettes Verständnis des Algorithmus.

Let's say that the doctor needs to have a certain amount of knowledge regarding the prediction that is on their hand and a certain amount of reasons to defend their belief regarding that prediction. They don't need to understand the AI as a system.

3.1.4 Auswirkungen des Algorithmus

Was sind die Auswirkungen der KI? Das ist eine Frage, die sich mehrere der Befragten stellten und auf die sich bei den meisten keine klare Antwort finden liess.

For me, the ethics points that are very important are understanding bias and limitations of systems in their impact. Sometimes it's hard to see how far that impact will go.

3.1.5 Qualitätsanforderungen und Sicherheit

Die Sicherheit der entwickelten Algorithmen zeigte sich als omnipräsentes Thema. Vordergründig ist die Patientensicherheit und die Auswirkungen auf die Gesellschaft.

The patient should not be harmed because of an AI system that might have a problem, an underlying problem that happened during the development of the system.

So that link between understanding that this is not just made-up data. This comes from real patients who have real suffering and have decided to help the research.

Das heisst, der direkte Schaden ist eigentlich erst vorhanden, wenn das Produkt an einem Menschen angewendet wird und es dieser Person Schaden zufügen kann, indem aufgrund einer KI eine falsche Entscheidung getroffen wird, die auf irgendeine Art zu einer schlechteren Gesundheitsversorgung führt.

Viele der Befragten sind sich dabei einig, was ein gutes und sicheres System ist: Ein zuverlässiges System, das besser ist als die vorherige Medizin, dem Patienten also nicht schaden soll und zu einer Verbesserung der Lebensqualität führt.

Das Problem ist, es wird nie hundertprozentig gut sein. Das ist etwas, was man nie erreichen wird. Es geht darum, überall den Durchschnitt anzuheben.

Für die Anwendung von KI-Systemen wird auch darauf hingewiesen, dass eine «Post-Market-Surveillance» helfen kann, Sicherheitslücken frühzeitig zu erkennen.

Hoffentlich hat das Produkt eine «Post-Market-Surveillance», also ein Monitoring, dass man, wenn etwas fehlschlägt, Prozesse hat, um herauszufinden, was passiert ist. Idealerweise wird man das früh genug auffangen, dass man reagieren kann.

3.1.6 Limitierungen

Wodurch ist ein System limitiert? Das ist eine Frage, die sich viele der Befragten stellen. Sie heben hervor, dass das Erkennen der Limitierungen und das eigene Bewusstsein derselben eine hohe Wichtigkeit haben. Das Kommunizieren der Grenzen des Algorithmus wird im *Abschnitt* 3.1.1 präsentiert. Von mehreren Befragten wurde die Datenart und -qualität als Limitierung angegeben.

> *So of course, the algorithms are only as good as the data that goes in and as broad as the data that goes in.*

Eine häufig erwähnte Limitierung von aktuellen KI-Algorithmen scheinen die noch nicht gut ausgeprägten Fähigkeiten zur Extrapolation[7] zu sein.

> *Machines today are really bad at extrapolating.*

Nicht die Nachfrage, sondern der Energieverbrauch von KI-Systemen wird als Limitierung wahrgenommen.

> *I don't think there's anything special, especially in medicine. In other sectors like, okay, OpenAI and ChatGPT and all these things. Of course, it's a different order of magnitude.*

3.1.7 Verantwortung

Wo liegt die Verantwortung und die Rechenschaftspflicht? Das ist eine Frage, die für die meisten der Expertinnen und Experten sehr schwierig zu beantworten war. So sind auch die Antworten auf die Frage, ob KI Verantwortung übernehmen kann und soll, sehr unterschiedlich ausgefallen.

> *I actually don't know the answer. It's difficult because we know that AI systems are not perfect. We know humans are not perfect.*

Die grundsätzliche Überlegung der Befragten war oft, ob die Verantwortung beim Menschen oder bei der Technik liegen kann oder muss.

> *Und deswegen würde ich sagen, wird auch in näherer Zukunft immer der Mensch im Mittelpunkt bleiben. Und die KI wird nicht eigenständig übernehmen.*

So auch das folgende Zitat:

> *Schlussendlich kann die KI keine Verantwortung übernehmen. Das ist ein Produkt.*

Die meisten der Befragten sehen keine Verantwortung auf Seiten der KI.

> *I don't have enough good reasons to believe that there are cases where the fully automated AI can do a good job.*

[7] Extrapolation bedeutet, dass ein System das Vermögen hat, Lösungen ausserhalb der Datenmenge zu finden und zu generieren. KI-Algorithmen sind zur Zeit nur begrenzt zur Extrapolation fähig. Interpolation ist im Gegensatz dazu, die Fähigkeit Muster innerhalb der bekannten Datenpunkte zu finden

Weiter folgte die Differenzierung der menschlichen Komponente in Ärzteschaft, Patienten sowie Unternehmen.

> *You have all these questions of liability. What happens if the AI system fails and who is liable? Is it the doctor or is it the company developing?*

Die Ärztin und der Arzt wurden in der Entscheidungsfindung hervorgehoben.

> *Nobody is going to be making a decision solely based on what an AI system is saying, at least from what I see.*

Die Entscheidungsfindung bleibt klar bei der Ärztin oder dem Arzt. Auf die Frage nach der Verantwortung wurde nicht eindeutig Stellung genommen.

> *Also die Verantwortung sollte nicht bei der KI liegen, aber es ist so, wenn man sich entscheidet…Also schlussendlich muss der Arzt entscheiden oder die Ärztin.*

In gewissen Fällen, so äusserte ein Befragter, könne KI Entscheidungen treffen. Die Frage nach der Verantwortlichkeit bleibt jedoch offen.

> *Ich glaube, wenn du ein KI-Modell hast, das hundertprozentig gut und besser als jeder Arzt ist, dann liegt die Entscheidung beim KI-Modell. Aber das Problem ist, wie wissen wir, dass das KI-Modell besser ist als jeder Arzt? Also denke ich, im Moment macht es absolut Sinn, dass diese bei den Ärzten liegt.*

Meistens wurde argumentiert, dass die Verantwortung schlussendlich beim behandelnden Arzt oder der behandelnden Ärztin liegt.

> *What is the doctor's role now in this world of AI? […] Obviously, you want that the MD at the end of the day signs off and has full control over the entire diagnosis.*

Auch der Patient wurde in der Frage nach Verantwortlichkeit erwähnt.

> *People are biased, but AI systems are biased as well. […] Probably at the end of the day, what needs to happen is a mode where the patient is informed about decisions from different sources.*

Einer der Befragten ordnete die Verantwortlichkeit dem Patienten selbst zu.

> *Maybe the patient is responsible at the end of the day. It's their body. It's their person. They are responsible. I don't know.*

3.1.8 Interdisziplinarität

Ein mehrfach erwähnter Aspekt ist die Interdisziplinarität. Für eine ethische Entwicklung eines KI-Systems sei die Zusammenarbeit zwischen verschiedenen Berufsgruppen mit unterschiedlichen Blickwinkeln wichtig. Es brauche einen Austausch mit Personen, die praktische Erfahrung, Fachwissen und das nötige Verständnis des medizinischen Hintergrunds haben. Ebenso scheint das Interesse an ethischer Weiterbildung vorhanden zu sein und auch der Wunsch nach Austausch mit Experten und Expertinnen aus

dem Bereich der Ethik. Insgesamt zeigte sich das Bedürfnis nach einer Verbesserung der Interdisziplinarität. In den folgenden Zitaten wird sichtbar, dass einige der Befragten der Meinung sind, die Entwicklung eines erfolgreichen KI-Systems benötige interdisziplinären Austausch.

> *I think that a successful AI is the one where you have this interdisciplinary conversation.*

Auch hier wurde auf die Notwendigkeit einer Zusammenarbeit hingewiesen.

> *Ich plädiere eigentlich immer für ein Tandem-Projekt.*

Mehrfach wurde das Bedürfnis nach ethischer Weiterbildung und nach mehr Unterstützung durch ethische Experten erwähnt.

> *I would like, actually, that my team would have more, let's say, become more sensitive to different kinds of ethical topics.*

Eine Befragte äusserte, dass sie selbständig an ethischen Weiterbildungen teilgenommen habe und sich für das Thema Ethik in der medizinischen KI-Anwendung und Implementierung interessiert.

> *I've taken a WHO class on ethics and AI that was recent…*

Aktuell ist die Zusammenarbeit mit Ethikerinnen und Ethikern noch nicht an der Tagesordnung, wie die folgenden zwei Zitate belegen.

> *No, I don't work with an ethicist on the team […]but maybe that should change.*

> *…wir haben nicht etwa die Situation, dass du mit einem ethischen Komitee arbeitest.*

Grundsätzlich wünschen sich die Befragten mehr Austausch mit Ethikerinnen und Ethikern. Ein Ethik-Mentoring wurde als Vorschlag genannt.

> *I think it could be improved. It could be improved that an ethics person could act as a devil's advocate during the development of the AI system.*

3.1.9 Datenhandhabung

Der ethische Umgang mit Daten stellte sich bei allen Befragten als ein zentrales Thema heraus, das in unterschiedlicher Weise immer wieder zur Sprache kam. Dies betraf Aspekte wie die Datenerhebung, -verarbeitung und -sicherheit, die Erkennung und Mitigierung von Bias sowie die Qualität, Quantität und Konsistenz der erzeugten Daten. Die meisten betonten, dass für sie Ethik in ihrem Alltag im sorgfältigen und bewussten Umgang mit Daten eine Rolle spiele.

Datenumgang Die folgenden Zitate demonstrieren, dass Datensicherheit und Datenschutz Themen sind, womit sich die meisten befassen. Verschiedene Aspekte im sicheren Umgang mit Daten wurden erwähnt. So zum Beispiel das Sichern und Ablegen von Daten.

…da musst du natürlich schon immer kurz überlegen, welche Daten du wo, wie, ablegst.

Alle Befragten waren sich darüber bewusst oder machten sich Gedanken dazu, dass Daten geschützt werden müssen, da es sich um persönliches und sensibles Material handelt.

Aber natürlich Datenschutz, das ist ein ganz, ganz wichtiger Aspekt, gerade wenn es um persönliche Daten geht. […] Das sind Informationen, die dürfen auf gar keinen Fall nach aussen gelangen, vor allem nicht an Arbeitgeber und Krankenkassen.

Ein Befragter äusserte, dass in diesem Bereich Forschungsprojekte getätigt werden, die eine Möglichkeit bieten, Daten möglichst real mit Hilfe von KI zu generieren:

Wenn wir mit Patientenbildern oder Patientendaten arbeiten möchten, dann müssen wir diese Daten schützen […] Und deswegen gibt es im Moment Forschungsprojekte, bei denen Modelle auf den Patientenbildern innerhalb eines Krankenhauses trainiert werden. Dieses Modell ist dann in der Lage, ähnliche Bilder zu generieren […] es generiert genau das gleiche Problem, zum Beispiel einen Tumor mit einer bestimmten Form, aber in einem leicht abgewandelten Körper, sodass diese nicht mehr zu einem Patienten rückverfolgt werden können.

Bias Bias wurden von den meisten Befragten als negativ bewertet wahrgenommen.

One of the big issues[…] is the source of the data that you're using in terms of demographics and any other biases that are implicit in the data. That's something that's always very, very important and relevant.

Die Befragten waren sich darüber einig, dass eine hohe Datenmenge und eine gute Datenqualität notwendig sind, um Bias zu minimieren.

We need a lot of work in order to generate synthetically, for example with LLMs, good data for rare diseases.

Zusätzlich wurde betont, wie wichtig der Prozess der Datensammlung ist. Die Patientengruppe, auf die der Algorithmus angewendet wird, muss im Datensatz widergespiegelt sein, damit möglichst adäquate Ergebnisse erzielt werden können.

There is a purely statistical aspect of bias, which is really important. If you have a bias that is expressed in a data set, like a certain type of patients is underrepresented in the data set. As a scientist, I believe that we should not think about, is this good or bad from a sociological perspective? We should think about this in terms of, did I sample my data correctly or not? First thing. Second, what does it mean in terms of applicability?

Für die Mehrheit hat das Wort Bias eine negative Bedeutung. Es ist für viele eine Datenverzerrung, die zu Nachteilen für spezifische Patientengruppen führen könnte.

We recently had something that was completely mind-blowing to me, which was about bias, and bias in AI. And that topic really, I mean, really affected me because there were some examples of bias that were so apparent that…but somehow you don't realize it in your day to day life.

And what I find really interesting was there is a statement made that we are not even aware of our own biases until AI reproduces our bias. Because if we are training algorithms and we ourselves have a bias, it will become known through that AI.

Es zeigte sich aber auch, dass der Begriff nicht für alle negativ konnotiert ist, sondern auch eine neutrale Bedeutung haben kann:

Bias is a term that I try to use in a neutral way. It's not positive or negative. It is something that exists.

Auch hier wurde eine relativ neutrale Position in Bezug auf Bias eingenommen:

And with bias, I mean, generally speaking, a tendency, a pattern. I don't know.

3.2 Sinn und Zweck des Algorithmus

Das genaue Definieren eines Ziels und die genaue Formulierung eines klinischen Endpunkts sind wichtige Voraussetzungen für eine ethische KI-Entwicklung, finden einige der Interviewpartnerinnen und Interviewpartner. Ein KI-System muss einen Nutzen haben, der festgelegt werden soll.

3.2.1 Nutzen und Notwendigkeit

In den folgenden Zitaten wird von den Befragten vor allem auf die Notwendigkeit eines Nutzens eingegangen.

These large language models are very interesting. Purely, again, from a functional point of view of what they can achieve, their utility in society, or at least the medical aspects of society, that for me is very open. […] I think these models will have a much bigger impact in society than they will have on medicine, for instance.

Sicherheit einerseits, aber andererseits auch in der Nutzbarkeit.

Der Wert des Algorithmus für einen bestimmten klinischen Endpunkt scheint wichtig zu sein:

I think now we're going to enter things where you really have to think a little bit more about what is the true value of an AI system for a clinical endpoint.

Eine Person äusserte dazu auch folgendermassen:

But I think in AI, and what I'm trying to push for is that the clinical goal should drive whatever AI learns.

3.2.2 Ziel des Algorithmus

Das Ziel soll genau formuliert werden und wie im *Abschnitt* 3.2.1 bereits erwähnt, einen genau definierten Nutzen haben. Alle Befragten waren sich einig, dass das Ziel des Algorithmus einen Benefit für Patienten, die Ärzteschaft oder die Gesellschaft bringen soll. Dies kann sich auf verschiedene Aspekte wie Gesundheit, Kosten oder

auch auf die Arbeitsbelastung beziehen. Sehr häufig wurde erwähnt, dass neue Systeme zu einer Entlastung der Ärztinnen und Ärzte und einer Effizienzsteigerung des Gesundheitssystems führen sollen.

> *Ein ganzes Hirn zu segmentieren und zu analysieren, dafür braucht der Arzt eine halbe Stunde. Ich habe auf meinem Computer zehn Sekunden.[…] Ich glaube, das ist etwas, was kommen muss, dass gewisse Prozesse schneller gehen müssen.*

Das bringt vor allem eine Zeitersparnis bei unterschiedlichen Aufgaben.

> *Ich definiere für mich eine erfolgreiche KI, wenn sie funktioniert. Und abgesehen davon, dass sie zusätzlich sehr genau arbeitet und auch viel Zeitersparnis einbringt.*

KI-Systeme sollen und können eine selektierende Rolle einnehmen, um unterschiedliche Prozesse zu beschleunigen.

> *[…] das Ganze schon mal ein bisschen vorab beschleunigen. Sachen, die kritischer sind, aussortieren, damit der Arzt sich dann darum kümmert und Sachen, die weniger kritisch sind, dass die dann durch die KI vielleicht in der Priorität tiefer eingestuft werden.*

Prozessoptimierungen und -beschleunigungen durch KI sollen dazu führen, dass die Arbeitsbelastung für die Ärztinnen und Ärzte sinkt und somit eine Entlastung eintreten kann.

> *AI-based, so deep learning algorithm that will help doctors. […] And so we thought, why not try to automate this process? […] And it should be a big relief, we hope, for our doctors here in the house.*

Generell wird die Arbeitsbelastung im Gesundheitswesen als Problem wahrgenommen, das mit Hilfe von KI gelöst werden könnte.

> *Und ich hoffe, dass der Arztberuf, wie er jetzt ist, mit der hohen Arbeitsbelastung ausstirbt und dafür ein mehr work-life-balance-freundlicher Beruf dadurch dann vielleicht zustande kommt.*

Die Beseitigung von sogenannten «dirty tasks» Aufgaben, die zeitraubend, mühsam, patientenfern und sehr bürokratisch sind, ist ein Ziel, das von mehreren Befragten hervorgehoben wurde.

> *It can do things much more rapidly. That was the thing because we're just counting things by eye. Instead of it being counted by AI, and you know that if you give the same image to two different people, they're going to count different things. For me, it was just an alleviation of a dirty task that we would otherwise have to do by eye.*

Im folgenden Beispiel wurde vor allem auf die administrativen Aufgaben eingegangen:

> *Dass Ärzte, die damit beschäftigt sind, administrative Aufgaben zu machen, zum Beispiel alte Patientennotizen einzuscannen…, das ist ja sehr stumpfe Arbeit. […] mit KIs werden wir einfach viel weniger Zeit mit manueller Arbeit verbringen, sondern mehr Zeit haben für andere Dinge.*

Dass die medizinischen KI-Systeme einen Vorteil für die Patienten bringen sollen, wurde sehr häufig erwähnt.

> *When you put together images and textual information and structured data, or at least information that is linked to that image, then it becomes a hugely powerful thing with so many advantages for medicine, research and education.*

KI soll dem medizinischen Personal helfen, bessere Entscheidungen zu treffen, was sich positiv auf den Patienten auswirken würde.

> *And I guess there are thousands of examples right now where AI is helping to give MDs a better choice, and patients as well.*

Ermächtigung des Patienten ist ein Argument, das mehrfach gefallen ist. KI-Systeme können sich dadurch profilieren.

> *It goes along with empowering the patient, which is ultimately what should probably be what happens.*

Etwa wie in folgendem Beispiel:

> *[…] to monitor patients at home.*

3.2.3 Zukunft in der KI-Entwicklung

Was Zukunftsvisionen sind und was Realität ist, scheint für die meisten klar zu sein. Allerdings ist in den Interviews aufgefallen, dass die Definition des Begriffs Künstliche Intelligenz unterschiedlich ausgefallen ist. Es wurde mehrmals erwähnt, dass KI keine «Magic Box» ist. Diese Formulierung zeigt eine rationale Anschauung der Künstlichen Intelligenz. Viele greifen auf den Begriff Algorithmus zurück, da dieser weniger vorbelastet ist.

Zukunftsvisionen KI hat sehr viel Potenzial und ist zugleich etwas, was viele Probleme lösen soll und kann. Zum aktuellen Zeitpunkt sehen die Befragten keine Gefahr oder Bedrohung durch KI für die menschliche Existenz. Bei allen ist die Stimmung tendenziell optimistisch und kaum von Angst geprägt.

> *Coupling smart instrumentation with AI systems to achieve things that we couldn't do before, that's something that I find really interesting.*

Tatsächlich machen sich einige Gedanken über eine Zukunft, wie sie in Science-Fiction-Filmen dargestellt wird.

> *…dann ist die Frage gar nicht unberechtigt: Wird KI tatsächlich irgendwann einmal selbstständig werden? Und dann uns Menschen als eine Gefahr für die eigene Existenz sehen? Aber das ist ein sehr weit hergeholtes Problem. Es ist nicht sehr nah. Es könnte eintreffen, aber davon sind wir noch weit entfernt, zum Glück.*

Oft ist die Realität so, dass noch viele Prozesse sehr grundlegend sind und die Zukunft darin liegt, diese in die klinische Anwendung zu bringen.

> *And now they're transitioning to products that hopefully can be integrated into the clinical realm, into real routine clinical use.*

KI als Magic Box KI besitzt keine magischen Kräfte und kann somit nichts aus Daten hervorzaubern. Das ist eine Meinung, die alle Befragten teilten.

> *Be aware that this is not magic and that the AI solutions are there to help.*

Diese Überzeugung zeigt sich auch in diesem Zitat sehr gut.

> *KI ist auch immer ein recht weiter Begriff und immer ein wenig eine Magic Box. In meinem Kopf ist KI einfach eine krasse Optimierung, die nicht viel anders ist als das menschliche Gehirn.*

KI wurde meistens pragmatisch als etwas Mathematisches beschrieben.

> *Es sind nur sehr schlaue Modelle oder sehr ausgeklügelte Modelle, die wir entwickelt haben, aber dahinter steckt im Endeffekt nichts weiter als Mathematik.*

Teilweise wird KI und das Erstellen dieser als etwas sehr Menschliches bezeichnet.

> *How we do AI, how we create these systems, in a sense, we leave a trace of ourselves in technology. I think that technology is probably the most human thing that we can do.*

3.3 KI als Werkzeug: Die Zukunft des Arztberufs

Inwieweit der Arztberuf aussterben könnte und welche Rolle die Ärztinnen und Ärzte in der Zukunft einnehmen, führt zu der Überlegung, ob künstliche Intelligenz letztlich alle ersetzen wird. Die Frage wurde im Wesentlichen damit beantwortet, dass KI zwar niemals die Ärztinnen und Ärzte vollständig ersetzen kann, jedoch immer mehr zu einem unverzichtbaren Instrument in der Praxis wird.

> *This is the terminator scenario where machine overruns the human and we are all done, and have to find a new job type of versus Ironman, where you as a doctor, you're empowered with this armor. That is technology that will help you. […] I don't think that doctors will become useless because AI systems today or probably in the next decade are seeing partial parts of the patient. These AI systems we developed are looking at pixels only.*

KI soll als unterstützendes Werkzeug dienen, als Rüstung, oder wie im folgenden Zitat erläutert wird, eine Art Engel sein, der hilft, den richtigen Entscheidungspfad einzuschlagen.

> *When you have to take a decision and you have the little devil and the little angel telling you things, the AI should be that little angel that says, okay, if you go that path with your patient, you should consider this.*

Warum eine KI Ärztinnen und Ärzte nicht ersetzen kann, begründen fast alle damit, dass ihr echte menschliche Werte, insbesondere Empathie, fehlen und diese auch durch blosse Nachahmung nicht kompensiert werden könne.

> *I don't want a machine to mimic that (empathy) because I know it's a computer program. An AI nurse doesn't work for me. I would feel cheated.*

Im folgenden Beispiel wird erwähnt, dass Empathie in gewissem Masse in KI einprogrammiert werden kann, diese aber das Bedürfnis nach wahrem menschlichem Mitgefühl nicht erfüllen würde.

> *I think that some of these aspects are almost already there. You say to ChatGPT, Oh, I'm having a bad day because of blah, blah, blah, and it will answer you back empathetically: Oh, I'm so sorry you are having a bad day.*

Auf diesen Unterschied macht der folgende Befragte aufmerksam:

> *Mitgefühl, würde ich sagen. Auf jeden Fall. Ich glaube, das ist immer der Hauptunterschied zwischen KI jetzt und dem Arzt.*

Gute Medizin erfordert viel mehr als Daten und Wissen.

> *No. […] it's not going to happen in the next, I don't know, a couple of generations easily because AI for the moment is still stupid. And medicine takes the whole …It's a holistic approach.*

Schlussendlich wird ein Aussterben des Arztberufes von allen verneint.

> *I think AI will just become part of the norm, but to replace that human aspect of medicine, that holistic aspect — no chance, not in the next few decades, maybe ever, hopefully.*

3.4 Standards und Richtlinien

Einerseits scheint der Wunsch und das Bedürfnis nach klaren ethischen Richtlinien und Standards präsent zu sein. Andererseits wurden die aktuellen Einschränkungen und das Vorgehen der Ethikkommission bei neuen KI-Produkten hinterfragt. Die Situation wurde von den Expertinnen und Experten so eingeschätzt, dass ein medizinisches KI-Produkt gleichwertig wie ein anderes Medizinalprodukt sein muss. Der Wunsch nach ethischen Richtlinien und ethischer Unterstützung wurde eher geäussert, als die Ablehnung gegenüber zu strengen Regulierungen.

3.4.1 Welche Rolle spielt die Ethikkommission?

Die Ethikkommission wurde mehrfach erwähnt und spielt im Berufsalltag aller Beteiligten eine Rolle. Sie wird jedoch als Hindernis wahrgenommen, nicht als ein Instrument, das eine ethische Implementierung der KI effektiv gewährleistet.

> *Everything related to ethics commissions and allowing for authorisation. But this has become so far from ethics that it's bureaucracy rather than ethics. I would call it more regulatory as opposed to ethics itself.*

Es wurde auch in Frage gestellt, ob die Ethikkommission über die notwendigen Fähigkeiten und Werkzeuge verfügt, um KI-Systeme angemessen zu beurteilen.

> *There, I sometimes also feel like they're taking on too much of a role, to be honest. They're not only maybe looking at the ethical aspects, but they're also looking at scientific aspects, which I don't know if they are in the capacity to be able to judge.*

3.4.2 Braucht es neue Ethikrichtlinien?

Die meisten liessen den Wunsch und das Bedürfnis nach klaren Hilfsmitteln laut werden, wie z.B. Richtlinien oder Standardprotokolle.

> *We don't have a standard protocol where we say, let's run the bias protocol or the bias detection protocol. We know, we read papers about those detection approaches. But I think that's something we have to improve in, in having more consistent bias evaluation or a bias framework to check our systems.*

Grundsätzlich sind sich aber fast alle einig, dass medizinische Produkte bereits sehr gut geschützt sind.

> *Although I genuinely believe that in the med tech space, in the space of building technology for medicine, we are already highly protected.*

Braucht es möglicherweise aussenstehende Instanzen, die Produkte ethisch zertifizieren? Das ist eine weitere Frage, die sich einige der Befragten stellten.

> *I don't think it's a bad idea because you know how things are. At the moment you are forced, it's officialized. Okay, now we have to do it.*

4 Diskussion

Was ist eine ethische KI und kann Ethik programmiert werden? Schon mit diesen Fragen öffnet man die Büchse der Pandora. Auf der Suche nach einer Antwort begegnet man unterschiedlichen und sehr komplexen Problemen. Um diesen auf den Grund zu gehen und einen Raum für Antworten zu schaffen, werden hier die Interviews mit Expertinnen und Experten in der medizinischen KI-Entwicklung diskutiert.

Eine Grundfrage, die bereits am Anfang der Suche für Probleme sorgt, ist die nach einer ethisch perfekten Welt und folglich einer ethisch perfekten KI. Wie kann eine perfekte Ethik überhaupt definiert werden und wer kann und soll das festlegen? Würden alle Menschen auf der ganzen Welt nach ihrer Meinung und Einstellung befragt werden, wäre das Ergebnis sehr wahrscheinlich ernüchternd – diskriminierend und ungerecht. Bei vielen Fragen wäre sich die Menschheit schon aufgrund der unterschiedlichen kulturellen und demographischen Gegebenheiten uneinig. Wie soll entschieden werden, was die gerechteste Lösung ist?

Mit dem Schleier des Nichtwissens (veil of ignorance) bietet der Philosoph John Rawls (1921-2002) in seiner Gerechtigkeitstheorie einen Lösungsansatz zur Findung einer gerechten medizinischen KI. Dieses moralphilosophische Gedankenexperiment hat das Ziel, einen möglichst gerechten Staat zu schaffen. Im Experiment wird Probanden aufgetragen, objektiv und unvoreingenommen Regeln und Gerechtigkeitsprinzipien für einen fiktiven Staat zu bestimmen. Die befragten Personen wissen nicht, welche Position sie in der fiktiven Welt einnehmen werden. Herkunft, Geschlecht, Hautfarbe, sozialer Status, geistige und physische Fähigkeiten etc. erfahren sie erst nach der Definition der Regeln. Das Unwissen über die zukünftige Rolle wird als Schleier des Nichtwissens bezeichnet.[28]

Folgendes Szenario veranschaulicht einige ethische Probleme beim Einsatz von KI im medizinischen Bereich: Eine KI hat das Ziel, möglichst viele Leben zu retten. Eine einfache Lösung ist es, einem gesunden Menschen das Leben zu nehmen, ihm alle Organe zu entnehmen und mehrere kranke Menschen damit zu retten. Ist diese Lösung ethisch? Die Mehrheit der Menschen würde diese Frage vermutlich verneinen. Anhand dieses Beispiels werden einige grundlegende Fragen sichtbar:

- Ist es richtig, möglichst viele Leben zu retten, in dem man aktiv den Tod von Wenigen in Kauf nimmt oder sogar bewirkt?
- Wer trifft Entscheidungen darüber, was in der Medizin richtig oder falsch ist?
- Nach welchen ethischen Normen soll eine medizinische KI programmiert sein?
- Was wenn eine medizinische KI nicht nach menschlichen Werten aligniert[8] ist?

Das «Trolley-Problem» und das Szenario der «Paperclip-Maximizing-AI» werfen

[8]Aligned (dt. ausgerichtet) von AI value alignment (dt. KI-Werteausrichtung) ist ein KI-System, das mit menschlichen Werten, Absichten und ethischen Standards im Einklang steht und diese fördert?

diese philosophischen Fragen auf und sollen als Ausgangspunkt der Diskussion dieser Arbeit dienen.

Das **Trolley-Problem** ist ein philosophisches Gedankenexperiment. Eine Person beobachtet, wie ein Zug auf fünf Personen zurollt. Sie kann die Weiche umstellen und dadurch verhindern, dass der Zug mit der Menschengruppe kollidiert. Auf dem Gleis, auf das der Zug umgeleitet wird, befindet sich nur eine Person. Ist es moralisch korrekt, den Tod einer Person durch eine aktive Handlung in Kauf zu nehmen, um den Tod der fünf anderen Personen zu verhindern?

Dieses Beispiel ist die Basis für die Diskussion der nachfolgenden Fragen, die auch in der Medizin relevant sind: Wie werden die Weichen gestellt und von wem? Entscheiden die Personen, die Algorithmen entwickeln, über richtig und falsch, also die Entwicklerinnen und Entwickler? Wer macht die Regeln und wie sollen diese lauten? Einig sind sich die Befragten darin, dass KI entsprechend unseren «menschlichen» Werten ausgerichtet sein soll.

Das Beispiel der **Paperclip-Maximizing-AI**, welche durch Nick Bostrom bereits 2003 vorgestellt und in Bostrom und Bostrom [29] ausführlich diskutiert wird, zeigt ein weiteres ethisches Problem, das durch die Anwendung von KI auftreten kann. Dieses Gedankenexperiment zeigt, was passieren könnte, wenn eine superintelligente KI das Ziel hat, möglichst viele Büroklammern herzustellen, jedoch nicht mit den ethischen Werten der Menschen abgestimmt ist. Eine solche KI würde ohne Rücksicht auf das Wohl der Menschen und deren Umwelt alles in Kauf nehmen, um ihr Ziel zu verwirklichen. Als Konsequenz davon würde die Welt mit Büroklammern überschwemmt, jegliches Metall zu Büroklammern verarbeitet und schlussendlich die Menschen beseitigt werden, um noch mehr Platz für noch mehr Büroklammern zu schaffen. Offensichtlich wäre diese Büroklammern-KI nicht gut mit den ethischen Werten der Menschen aligniert. Dieses Beispiel zeigt, was möglicherweise passieren könnte, wenn eine KI nur den Auftrag bekäme, möglichst viele Menschenleben zu retten.

Die Wichtigkeit der richtigen Wertausrichtung wird in einem für das World Economic Forum 2024 veröffentlichten Artikel betont:

> *Value alignment is linked to the concept of AI ethics red lines – the non-negotiable boundaries that AI systems must not cross. By embedding core human values and maintaining rigorous oversight, the value alignment process makes sure AI systems operate within established moral and legal frameworks, safeguarding against unethical behaviour and maintaining societal trust.*[30]

Da Ethik eine individuelle und keine universelle Angelegenheit und von vielen soziokulturellen Faktoren abhängig ist, braucht es eine gewisse Hilfeleistung, um ein KI-System zu entwickeln, das mit den menschlichen Werten aligniert ist. Eine solche Hilfeleistung wurde auch mehrfach von den Entwicklerinnen und Entwicklern gefordert. Für eine ethische KI-Entwicklung können die vier medizinischen Grundprinzipien von Beauchamp und Childress[6], die FMH Standesordnung[23] oder auch die SAMW medizinethischen Richtlinien herangezogen werden. Zusätzlich können sowohl staatliche Gesetze und Normen als auch die Ethikkommission behilflich sein.

4.1 Datenhandhabung und Umgang mit Bias

«KI Modell empfiehlt: Appenzeller Alpenbitter ist das wirksamste Grippemittel!» Dieses Ergebnis könnte sich herausstellen, wenn eine KI mit Daten aus einer nicht repräsentativen Befragung in einer Appenzeller Landarztpraxis zur Wirksamkeit von Grippemedikamenten trainiert wurde. Würde dieses Modell schweizweit zum Einsatz kommen, wäre die Therapieempfehlung bei jedem Patienten mit einer Grippeinfektion die Einnahme von Appenzeller Alpenbitter.

Die Güte, Zuverlässigkeit und Leistungsfähigkeit eines Algorithmus hängen, wie in dem Beispiel auch in der Medizin, von der Qualität und Vielfalt der Daten ab, die dafür verwendet werden. Ohne hochwertige und repräsentative Daten ist ein Algorithmus wenig wert.

Die befragten Expertinnen und Experten aus dem Bereich der KI-Entwicklung sind ebenfalls dieser Meinung: Ein sorgfältiger Datenumgang ist zentral und wurde häufig angesprochen. Die meisten Entwicklerinnen und Entwickler sind sich bewusst, dass die Leistung ihres Systems stark abhängig von den Daten und deren Bearbeitung ist. Zur Frage, wie ein sorgfältiger Umgang gewährleistet werden kann, gibt es unterschiedliche Ideen und Meinungen, wie eine gute Schulung, Leitlinien oder interdisziplinäre Zusammenarbeit mit Ethikerinnen und Ethikern. Ausserdem wurde die Idee aufgegriffen, dass ein Institut oder Unternehmen, das Forschung betreibt, Kohorten von Patientendaten sorgfältig sammeln und kuratieren soll, um damit Forschung zu betreiben und Produkte zu entwickeln. Die Trainingsdaten von Algorithmen stammen von Patienten; Menschen, die ein Leiden haben, krank oder verstorben sind. Sie haben ihre persönlichen Daten der Forschung zur Verfügung gestellt und da diese Daten den Patienten angehören, ist ein besonders sensibler Umgang notwendig. Es zeigten sich viele der Befragten besorgt, dass zu wenig Sensibilität bei der Datennutzung und -verarbeitung vorhanden sei.

Dazu schreiben Mainzer und Kahle [19]: «...es sollte aber zudem klar sein, dass man mit Hilfe maschinellen Lernens, wenn man von schlechten Daten ausgeht, kaum gute Resultate erwarten kann.» Ergänzend wurde darauf hingewiesen, dass nicht nur die Qualität der Daten ausreichend sein soll, sondern auch die Qualität der Labels, mit denen die Daten versehen sind. Eine sorgfältige Auswahl der Labels ist notwendig und kann dazu beitragen, Bias und damit Risiken und Ungerechtigkeiten zu minimieren und doch von den Vorteilen einer KI zu profitieren.[31]

Die Expertinnen und Experten messen Bias eine wichtige Rolle in der Medizin und in der Wissenschaft bei, aber es werde im Alltag zu wenig darüber gesprochen. Ein komplettes Bewusstsein über die eigenen Vorurteile zu erlangen, ist schwierig, wenn nicht sogar unmöglich. KI soll nicht nur als Gefahr einer Verstärkung dieser Vorurteile und Bias gesehen werden, sondern auch als Chance, uns diese vor Augen zu führen und sie bestmöglich zu mitigieren. Einige Expertinnen und Experten sind dabei der Meinung, dass Bias per se keine Gefahr darstellen, sondern ein unreflektierter und nachlässiger Umgang damit. Durch eine kritische Auseinandersetzung mit den gesammelten Patientendaten verliert der Begriff Bias auch seine negative Konnotation und wird zu etwas Neutralem, einem Begriff ohne positive oder negative Wertigkeit.

Bias entstehen auf mehreren Ebenen. Bereits die Daten, die zum Training einer KI verwendet werden, können zu Bias führen. Bias können durch das Design von KI-Systemen beeinflusst und verstärkt werden. Die Anwendung, sowohl der Ärztinnen und Ärzte als auch der Patienten, hat zudem einen Einfluss auf das Ergebnis.[32]

«What can cause a Machine-Learning System to be unfair?» – Mit dieser Frage beschäftigen sich Rajkomar, Hardt, Howell, Corrado und Chin [33]. Sie betrachten die Fairness eines Algorithmus auf mehreren Ebenen und zeigen, dass sich diese unter anderem in verschiedenen Formen von Bias widerspiegeln kann. So entstehen etwa sogenannte «Label Bias», wenn die Trainingsdaten gesellschaftliche oder gesundheitssystembedingte Ungleichheiten enthalten – beispielsweise, wenn bestimmte Gruppen systematisch falsch diagnostiziert wurden. «Minority Bias» treten auf, wenn bestimmte Bevölkerungsgruppen in den Trainingsdaten unterrepräsentiert sind. Wird der Output eines KI-Systems unkritisch übernommen und blind darauf vertraut, spricht man von «Automation Bias». In ähnlicher Weise beschreibt der Begriff «Dismissal Bias» Situationen, in denen Warnungen oder Hinweise des Systems ignoriert werden. Werden solche Verzerrungen, sei es im Design, in den Daten oder in der Implementierung, nicht berücksichtigt, kann dies dazu führen, dass KI-Systeme gesundheitliche Ungleichheiten nicht nur reproduzieren, sondern sogar exazerbieren.

Fairness sollte daher ein zentraler Aspekt des Entwicklungsprozesses einer KI sein. Damit wird gefördert, dass möglichst alle Patienten von einer KI-Technologie profitieren können und nicht aufgrund von Bias benachteiligt werden. Eine faire KI beinhaltet verschiedene Kernpunkte, die in *Tabelle* 3 dargestellt sind.

Equal Outcomes (dt. gleiche Ergebnisse)	**Equal Performance (dt. gleiche Leistung)**	**Equal Allocation (dt. gleiche Allokation)**	**Trade-offs (dt. Abwägungen)**
Gleiche Ergebnisse zu generieren bedeutet, dass Ergebnisse in allen Patientengruppen ähnliche Vorteile entstehen. Zudem kann in einer stärkeren Form noch dazu beigetragen werden, dass Ungleichheiten durch die Ergebnisse sogar verringert werden.	Ein Modell sollte in allen Patientengruppen gleich genaue Ergebnisse und Leistungen erbringen. Die Sensitivität, Spezifität und der positiv prädiktive Wert definieren die Leistung eines Systems und sollen immer gleich gut sein.	Eine gleiche Allokation bedeutet, dass Ressourcen so auf die Patienten verteilt werden, wie sie nötig sind und beispielsweise nicht nach den finanziellen Möglichkeiten einer Person.	Für eine möglichst faire Lösung gibt es Situationen, in denen ethische und kontextbezogene Abwägungen der ersten drei Punkte erfolgen müssen. Werden diese Punkte strikt befolgt, müssen nicht zwingend faire Lösungen entstehen. Ein Abwägen dieser Punkte ist manchmal notwendig.

Tabelle 3: Kernpunkte einer fairen KI-Entwicklung

Sämtliche Produkte in der Medizin haben einen direkten Einfluss auf die Gesundheit von Menschen. Eine sorgfältige und kontrollierte Entwicklung sollte also nicht nur bei Medikamenten gewährleistet sein, sondern auch bei KI-Produkten. Gemäss Swissmedic sind medizinische Produkte, so auch die Datenhandhabung und die Datenakquisition, den folgenden Gesetzen und Verordnungen unterstellt: Dem Heilmittelgesetz (HMG), der Medizinproduktverordnung (MepV), der Verordnung über In-vitro-Diagnostika (IvDV), dem Humanforschungsgesetz und der Verordnung über klinische Versuche mit Medizinprodukten (KlinV-Mep). Im Gegensatz zu Arzneimitteln gilt für Medizinprodukte keine behördliche Zulassung.[34] Ziel dieser Richtlinien und Regulationen ist es, den Schutz von Leib und Leben der Patienten gewährleisten zu können.

Fakt ist, dass momentan nicht alle Menschen komplett in Datenpunkten erfasst werden können. Auch wenn das irgendwann möglich sein könnte, bleibt die Anwendung der Daten eine Herausforderung. Eine Lösung für eine gerechte KI wäre damit noch nicht gefunden.

Im Alltag der Softwareentwicklerinnen und -entwickler resultieren Schwierigkeiten oftmals aus sehr praxisnahen und grundlegenden Ursachen. Ein Beispiel dafür ist eine Erkennungssoftware für Myokardinfarkte. Wenn diese bei Männern häufiger einen Myokardinfarkt als bei Frauen diagnostiziert, wirft diese Tatsache Fragen auf: Kommt es tatsächlich bei Männern häufiger zu Myokardinfarkten? Basiert die Software primär auf Daten von Männern, die einen Myokardinfarkt erlitten haben? Gibt es Bias in der Datenerhebung oder liegen Softwarefehler vor?

4.2 Nachvollziehbarkeit und Transparenz: Das Black Box Dilemma

Das Thema Black Box wurde von den Experten rege aufgegriffen. Alles sei eine Black Box, es komme nur auf den Betrachter an, ist eine Aussage aus den Gesprächen. Sie selbst können meist nicht alles nachvollziehen, was eine KI generiert und weshalb. Ein neuronales Netzwerk aus vielen Schichten, die wiederum auf vielen Datenpunkten basieren, generiert Lösungen, die nicht mehr 1:1 nachvollziehbar sind.

In den Gesprächen zeigten sich mehrere Faktoren, die die Nachvollziehbarkeit von KI erschweren. Einerseits sei es wichtig, dass man sowohl wissen müsse, nach was zu suchen sei, als auch wie zu suchen sei. Die Black Box wird als Kiste beschrieben, in die zwar hineingeschaut werden könne, es jedoch sehr schwierig sei, alles zu verstehen, was die KI generiere, da die Komplexität eines «Gedankengangs» der KI enorm riesig sei. Andererseits seien die vielen Dimensionen ein weiterer verkomplizierender Faktor. Ein Experte wies darauf hin, dass nicht bekannt sei, in wie vielen Dimensionen eine KI agieren könne. Für uns Menschen ist eine Welt in drei Dimensionen üblich. Wird die Zeit als vierte Dimension dazugerechnet, fällt es dem menschlichen Vorstellungsvermögen bereits schwieriger, ein System zu verstehen. Besteht ein System aus mehr als vier Dimensionen, wird es für das menschliche Hirn fast unmöglich, diese noch verstehen und nachvollziehen zu können.

Muss ein Tool oder eine Maschine komplett verstanden werden? Eine Position, die in den Gesprächen vertreten wurde, lautet: Nein, die meisten Dinge, die wir im Alltag benützen, werden von den Anwendern nicht verstanden. Nicht nur digitale Geräte sind in ihrer Funktion nicht mehr komplett nachvollziehbar, sondern auch viele andere Technologien im Alltag. Das Beispiel, das in den Interviews genannt wurde, sind Smartphones. Praktisch alle benutzen sie so rege, dass ein Leben ohne sie nicht mehr vorstellbar ist. Doch kaum ein Mensch, geschweige denn ein Arzt oder eine Ärztin, wird komplett nachvollziehen können, wie ein Smartphone funktioniert.

> *Machine learning algorithms and neural networks are complex to the point that their internal workings are not straightforward to understand, even for subject experts. While they remain purely technical and determined systems, it is impossible (partly because they are learning systems and therefore change) to fully understand their internal working.*[35]

Das menschliche Hirn selbst ist eine Black Box. Es besteht aus 86 Milliarden Neuronen, deren Netzwerk und kausales Lernen bis jetzt noch nicht ganz verstanden werden, wie Mainzer und Kahle [19] schreiben. Wir verstehen also im Grunde genommen nicht einmal unsere eigene Software komplett.

Ärztinnen und Ärzte sind es in der Regel gewohnt, alles verstehen zu müssen und zu wollen, was sie tun. Es scheint die Erwartung der Medizinerinnen und Mediziner zu sein, ein neues KI-Tool verstehen zu müssen. Bei genauer Betrachtung werden schon jetzt viele Geräte, Maschinen und Software benützt, die nicht mehr vollständig verstanden werden. Beispiele sind MRI-Geräte oder PET-CTs sein, die für die meisten Ärztinnen und Ärzte wie Black Boxes sind, oder auch die Wirkungsweisen von Medikamenten, die noch nicht ausreichend erklärt werden können. Dazu schreibt Eric Topol in seinem Buch

«Deep Medicine»: Viele Therapien in der Medizin, wie Elektrokonvulsionstherapien bei Depressionen oder Wirkungsweisen von gewissen Medikamenten, sind auch noch nicht komplett verstanden.[17]

Die meisten Entwicklerinnen und Entwickler erwarteten in der Regel ein gewisses Wissen der Ärztinnen und Ärzte gegenüber den Voraussagen und Funktionsweisen einer KI, jedoch nicht ein komplettes Verstehen des Algorithmus. Für die Entwicklerinnen und Entwickler scheint die Erklärbarkeit nicht immer eine generelle Anforderung zu sein. Ein Vorschlag eines Experten war es, einen Konversationsmodus, also einen Chat-Bot für Ärztinnen und Ärzte einzubauen, der Fragen zur Lösung beantworten könnte, wie: «Ich habe das Gefühl, es gäbe eine bessere Variante für dieses Problem. Zeig mir die Fakten und den Lösungsweg dazu, weil ich (als Arzt) nicht gleicher Meinung bin.» So könnte eine Diskussion mit der KI stattfinden, wie sie auch unter Fachpersonen im Gesundheitswesen jeden Tag stattfindet.

Ganz anders klingt es im durch die Europäische Union 2024 veröffentlichten EU AI Act: «Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights.» Und weiter: «Requirements should apply to high-risk AI systems as regards risk management, the quality and relevance of data sets used, technical documentation and record-keeping, transparency and the provision of information to deployers, human oversight, and robustness, accuracy and cybersecurity. Those requirements are necessary to effectively mitigate the risks for health, safety and fundamental rights. As no other less trade restrictive measures are reasonably available those requirements are not unjustified restrictions to trade.» Der EU AI Act scheint in Bezug auf die Nachvollziehbarkeit wenig Spielraum offen zu lassen.[9]

Regina Barzilay MIT School of Engineering Distinguished Professor of AI and Health ist in Knight [36] der Meinung, dass KI ein riesiges Potenzial hat, die Medizin zu revolutionieren. Sie hat jedoch auch das Gefühl, dass es notwendig sei, dass KI-Systeme in der Lage sein sollten, Entscheidungen oder Schlussfolgerungen für Ärztinnen und Ärzte nachvollziehbar zu machen und diese zu erklären. Es sei wichtig, dass Maschinen und Menschen interagieren können.

Wir Menschen stellen die Frage «Warum?» oder «Wieso?», wenn wir etwas nicht verstehen und genauer erfahren möchten. Eine KI kann dies insofern beantworten, wie sie das statistisch gelernt hat. Sie müsste also eine dem ärztlichen Hirn verständliche Antwort liefern und nicht nur ein: «Weil ich es so gelernt habe.» wie Mainzer und Kahle [19] erläutern. Dies würde zwar die Black Box nicht komplett enthüllen, jedoch eine dem Menschen nachvollziehbare Antwort liefern.

Die Prinzipien der Medizinethik nach Beauchamp und Childress [6] konkurrieren nur teilweise mit einer Black Box: Die Autonomie eines Patienten kann dadurch eingeschränkt sein, wenn beispielsweise eine Therapieempfehlung einer KI nicht komplett transparent ist und der Patient in seiner Entscheidungsfindung zumindest teilweise bevormundet wird. Das Prinzip des Nichtschadens kann verletzt werden, wenn falsche Diagnosen auf Basis von fehlerhaften Entscheidungen einer KI entstanden sind und es

aufgrund einer ungenügenden Transparenz zu einer erschwerten Suche der Fehlerquelle kommt.

Zusammenfassend kann gesagt werden, dass es grundsätzlich zwei unterschiedliche Positionen gibt. Aus Sicht der Entwicklerinnen und Entwickler wird eine gewisse Opazität in Kauf genommen, solange eine gute Funktionalität gewährleistet werden kann. Den Gegenpol bilden die gesetzgebenden und regulierenden Behörden. Diese fordern eine «glass box», also eine komplett nachvollziehbare KI. Wird akzeptiert, dass die KI nicht transparent ist, vereinfacht das die Integration in den medizinischen Alltag signifikant. Die Frage nach der Verantwortung stellt sich aber umso brennender.

4.3 Verantwortlichkeit und Entscheidung bei der Nutzung von KI in der Medizin

Neue und komplexe Fragen der Verantwortung und Haftung entstehen bei der Integration von KI im Gesundheitswesen. Bleibt bei der erfolgreichen Anwendung von KI oft unklar, wem die Lorbeeren gebühren, so wird die Frage der Verantwortung bei Fehlern umso dringlicher. Wer ist verantwortlich und wer haftet, wenn eine KI zu einer Fehldiagnose oder Fehlbehandlung führt?

4.3.1 Verantwortlichkeit in der Medizin

Was ist mit Verantwortung gemeint? Gemäss Mainzer und Kahle [19] ist Verantwortung juristisch gesehen die Pflicht einer Person, für ihre Entscheidungen und Handlungen nach geltendem Gesetz Rechenschaft ablegen zu müssen. Diese juristische Verantwortung ist also genau genommen nicht eine moralische oder religiöse Verantwortung (Gewissen).

Die Verantwortung gegenüber Patienten unterscheidet sich von der gesetzlichen Verantwortlichkeit. Verantwortung umfasst nicht allein die Zuordnung von Haftung, kann auch eine ethische Verpflichtung gegenüber Umwelt, Ressourcen und Gesellschaft einschliessen. Das Handeln einer KI oder eines Mediziners oder einer Medizinerin könnte etwa als verantwortungslos gelten, ohne dass es rechtliche Konsequenzen hätte. Folglich ist es möglich, dass in bestimmten Fällen keine juristische Haftung besteht, wenngleich die moralische Verantwortung des Arztes oder der Ärztin fortbesteht. Es scheint, dass das Gesundheitswesen nicht nur den Patienten gegenüber Verantwortung tragen soll, sondern auch gegenüber der Umwelt. Diese Ansicht wurde auch von einigen Entwicklerinnen und Entwicklern geteilt.

AlphaGo Zero von Google hat beispielsweise 96 Tonnen CO_2 innerhalb von 40 Tagen generiert, was ungefähr 1000 Flugstunden entspricht. Das wirft die Frage auf, ob der Energieverbrauch von KI gerechtfertigt ist. Nur weil sich eine KI an geltende Gesetze hält, muss sie noch lange nicht moralisch sein, schreibt Wynsberghe [37].

Positive Verantwortung und Anerkennung Solange Programme und Systeme funktionieren, Erfolge erzielen und gute Ergebnisse liefern, möchten alle am Erfolg teilhaben. Dennoch ist hier nicht klar, wer oder was hier die Leistung vollbracht hat. Wem die Anerkennung zusteht und somit die positive Verantwortung bleibt unklar.

Sind es die Softwareentwicklerinnen und Softwareentwickler, das Unternehmen oder gar die KI selbst? Die Zuweisung dieser positiven Verantwortung bleibt oft unklar. Eine ausbleibende Anerkennung ist meist jedoch weniger problematisch als eine ausbleibende Bestrafung.

Negative Verantwortung und Verantwortungslücken Interessanter wird es, wenn ein System nicht funktioniert. Wenn es im Falle einer medizinischen KI zu einer Fehldiagnose, einer verpassten Diagnose oder gar zum Tod eines Patienten kommt - wer oder was ist dann dafür verantwortlich?

Der Fall eines selbstfahrenden, KI-basierten Uber-Autos, das 2018 bei einer Testfahrt in Arizona eine Frau tödlich verletzte, schlug starke Wellen. Wer war nun verantwortlich für diesen tragischen Unfall? Uber als Unternehmen, das Auto selbst oder die Entwicklerinnen und Entwickler?[8] Ein ähnlicher Vorfall ereignete sich auch 2016 in den USA, als durch ein teilweise autonomes Auto der Marke Tesla der Fahrer beim Unfall getötet wurde. Die NZZ schrieb zu diesem Ereignis:

> *Der Vorfall wirft die Frage auf, inwiefern sich Mensch und Maschine Verantwortung teilen können.*[38]

Stellt man ChatGPT die Frage: «Wer ist verantwortlich wenn du einen Fehler machst?», gibt die KI folgende Antwort:

> *Wenn ich (als KI-Modell) einen Fehler mache, ist die Verantwortung eine komplexe Frage, da ich selbst weder moralische noch rechtliche Verantwortung tragen kann. Stattdessen liegt die Verantwortung bei den Menschen und Organisationen, die mich entwickelt, trainiert und bereitgestellt haben. Hier eine Aufschlüsselung:*
>
> - *OpenAI (Entwickler der KI)*
> - *Nutzer:Innen*
> - *Plattformanbieter*[39]

Wie Chat-GPT bereits schon selbst beantwortet hat, kann KI momentan keine Verantwortung übernehmen. Durch den Einsatz von komplexen Technologien wie KI wird es jedoch zunehmend komplexer, moralische oder rechtliche Schuld jemandem zuzuweisen und zur Verantwortung zu ziehen. Bleibt aufgrund einer Fehlfunktion einer KI unklar, wer für den Fehler verantwortlich ist, wird von einer «Verantwortungslücke» (engl. responsibility gap) gesprochen und es kann niemand haftbar gemacht werden. Dies tritt beispielsweise auf, wenn autonome Systeme Entscheidungen treffen, deren Entscheidungen weder vollständig vorhersehbar noch kontrollierbar sind.[40] Diese Verantwortungslücken werden in der Medizin besonders interessant, weil hier direkt über Menschen und Menschenleben entschieden wird.

4.3.2 Stakeholder der Entscheidungsfindung in der Medizin

In diesem Abschnitt wird untersucht, welche Stakeholder in die Entscheidungsprozesse im Gesundheitssektor involviert sind, um die Verantwortungsfrage genauer zu analysieren. Welche Entitäten sind an medizinischen Entscheidungen beteiligt, wenn eine KI im Gesundheitswesen eingesetzt wird?

Folgende Stakeholder in der *Abbildung* 2 spielen im Gesundheitswesen eine Rolle:

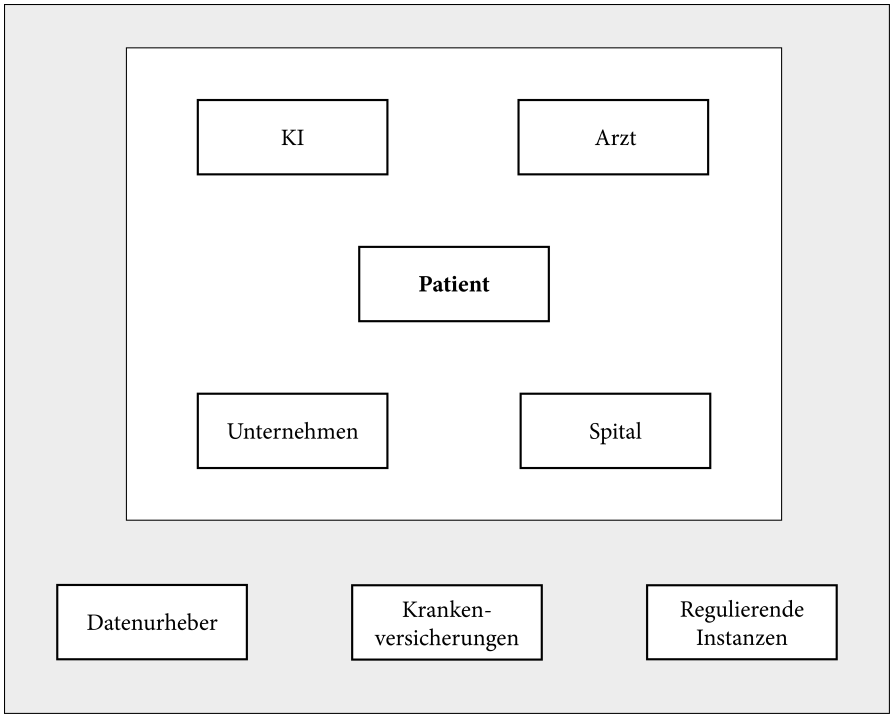

Abbildung 2: Stakeholder im Gesundheitswesen

Aufgrund dieser Aufteilung kann die Verantwortungsfrage in den weiteren Abschnitten besser diskutiert werden.

4.3.3 Verantwortlichkeitszuschreibungen

In den Interviews mit den Entwicklerinnen und Entwicklern traten verschiedene Perspektiven bezüglich der Verantwortungszuweisung zutage. Diese variierte von: «Der Arzt ist verantwortlich.» über «Der Patient selbst ist verantwortlich.» bis zu «Die Unternehmen sind verantwortlich.» In den Expertengesprächen wurde insbesondere auf zwei verschiedene Szenarien eingegangen:

A) Die KI wird lediglich als Werkzeug verwendet und nimmt eine Hilfsrolle bei der Entscheidungsfindung ein. Das könnte zum Beispiel eine KI zur Erkennung von Lungenkrebs in CT-Bildern sein. Der Arzt oder die Ärztin können das Ergebnis selbständig begutachten und dann aufgrund dieses Wissens eine Entscheidung zur definitiven Diagnose und Behandlung treffen.

B) Die KI hat alle Daten wie Bilddaten, Behandlungsschemata und individuelle Patientendaten zur Verfügung. Sie kann selbständig eine genaue Diagnose und Therapie definieren und initiieren.

Je nach Szenario wurden durch die Entwicklerinnen und Entwickler verschiedene Zuordnungen der Verantwortung vorgenommen.

Einige Entwicklerinnen und Entwickler sind der Meinung, dass KI primär als Werkzeug (Variante A) fungieren und folglich die Verantwortung weiterhin beim Arzt oder der Ärztin liegen wird. Dieser sollte stets die Entscheidungsgewalt und die Kontrolle über die Behandlung eines Patienten behalten. Zusätzlich wird argumentiert, dass KI immer ein Produkt sei, und ein Produkt könne keine Verantwortung übernehmen. Die meisten Befragten sehen deshalb keine Verantwortung auf Seiten der KI.
Lange waren Computer nicht in der Lage, medizinische Entscheidungen selbständig zu treffen. Mit der Anwendung von KI verändert sich der Status quo insofern, dass Maschinen autonom zu Lösungsansätzen kommen und alleine Entscheidungen treffen können. Die Tatsache, dass lediglich der Arzt oder die Ärztin und der Patient am Entscheidungsprozess beteiligt sind, ist plötzlich nicht mehr gegeben. Die Europäische Datenschutz-Grundverordnung (DSGVO) besagt, dass der Nutzer einer KI, ob Unternehmen (Krankenhaus oder Praxis) oder Einzelperson (Arzt/Ärztin), primär verantwortlich für eine Fehlfunktion ist. Hersteller können jedoch zusätzlich zur Verantwortung gezogen werden.[41] Folglich weist die DSGVO tendenziell dem Nutzer die Verantwortung zu. Erfüllen jedoch alle Akteure ihre Verhaltenspflicht, kann nach bisherigem geltendem europäischen Recht bei einem Maschinenversagen keine Person haftbar gemacht werden.[42] Dann bleibt eine Verantwortungslücke übrig.

Sollten KI-Algorithmen jedoch ein Niveau erreichen, bei dem eine direkte Kontrolle nicht mehr möglich ist und ihre Resultate unvorhersehbar werden (Variante B), sei es ungerecht, Menschen zu bestrafen, wenn diese Technologien Schaden verursachen oder tödliche Folgen haben, argumentiert der Philosoph Sven Nyholm. Es sei problematisch, wenn Menschen die Verantwortung tragen müssen, um Verantwortungslücken zu füllen, die durch KI-Technologien entstehen. Ein Gespann, bestehend aus Mensch-Maschine, in Betracht zu ziehen, wäre ein Vorschlag.[8]
Inwiefern das jedoch möglich ist, eine Verantwortung überhaupt auf eine Maschine, sei es nur teilweise, zu übertragen, ist fraglich. Die Verantwortung wäre, zumindest zum jetzigen Zeitpunkt, dennoch vollständig beim Menschen. Der Vorschlag von Nyholm, von einer gemeinsamen Verantwortung aus Mensch und Maschine, ist durchaus interessant. Die Grundfrage, ob eine Maschine überhaupt Verantwortung tragen kann, bleibt offen. Eine Maschine ohne Gewissen, ohne Emotionen, ohne Bewusstsein, könnte zwar formal bestraft werden, die Strafe hätte aber nicht wirklich einen Effekt. Gemäss den derzeitigen rechtlichen Bestimmungen in der Schweiz kann eine Maschine nicht

strafrechtlich zur Verantwortung gezogen werden. Für eine Maschine wäre eine Gefängnisstrafe vermutlich keine Strafe, vorausgesetzt sie besitzt nicht die Fähigkeit zu leiden. Einer Maschine wäre eine Gefängnisstrafe vermutlich ziemlich egal, eine Todesstrafe sowieso und eine Geldstrafe könnte sie nicht bezahlen. Selbst wenn Maschinen eine Form von Verantwortung zugeteilt würde, hätte dies vorerst keine Konsequenzen, zumindest solange Maschinen kein Bewusstsein und keine Emotionen besitzen.

Als weitere mögliche Verantwortliche kommen Unternehmen in Frage. Den Unternehmen die Verantwortung zuzuweisen, wurde in den Interviews mehrfach diskutiert.

Unternehmen spielen eine zentrale Rolle in der Verantwortlichkeitszuweisung. Ist ein Produkt auf dem Markt fehlerhaft, haftet meist das Unternehmen dafür. Das ist Normalität - auch in der Medizin. Bei einem Fehler, der Auswirkungen auf Patienten und Menschenleben hat, sollen diese zumindest teilweise dafür verantwortlich sein. Warum soll sich der Zustand bei KI-Produkten ändern? Die klaren Grenzen zwischen Produkt, Unternehmen und Anwender verändern sich. Plötzlich entstehen Produkte, die selber Entscheidungen treffen können, auf die das Unternehmen und die Anwender keinen Einfluss mehr haben.

Dazu äussert sich der Charakter Stan in der sozialkritischen Serie «South Park», in einer Folge, in der es um den Aufstieg von generativer KI und den wachsenden Einfluss von grossen Techunternehmen geht, folgendermassen:

> *Look everyone. We can't blame people that are using ChatGPT; it's not their fault. I'll tell you whose fault it is. It's the giant tech companies who took OpenAI, packaged it, monetized it, and pushed it out to all of us as fast as they could in order to get ahead. OpenAI is so powerful, that it has to be something that everyone can use, control, and contribute to or else AI will be controlled by corporations that just want an unfair advantage like Cartman does.*[43]

Ist es ungerecht, ein Unternehmen, das einen KI-Operationsroboter entwickelt hat, der besser operiert, als dies der beste Chirurg oder die beste Chirurgin gekonnt hätte, zu bestrafen, während der Chirurg oder die Chirurgin unter dem Strich weniger Leben gerettet hätte als die KI. Diese Argumentation überzeugt aber insofern nicht, weil schon immer neue Technologien auf den Markt kamen, die etwas besser konnten als der Mensch. Wenn Menschen potenziell zu Schaden kommen könnten, muss die Verantwortlichkeit immer wieder aufs Neue geklärt werden, auch wenn sie die Expertise der besten Menschen überschreitet. Fortschritt und Entwicklung gehören zum Menschen. Technologien müssen immer wieder als neuen Status Quo angenommen, an den Recht und Gesetz angepasst werden.

Eine in den Interviews selten vertretene Auffassung ist, dass auch Patienten eine gewisse Verantwortung tragen könnten. Bedingung hierfür sei, dass Patienten umfassend und akkurat informiert werden, beispielsweise durch den sogenannten Informed Consent. Informed Consent (informierte Einwilligung) ist ein Konzept aus der Medizin und aus dem Humanforschungskontext. Er besagt, dass Patienten oder Probanden ausreichend und verständlich aufgeklärt und informiert werden müssen, damit eine eigene Entscheidung getroffen werden kann. Der Informed Consent führt damit zur

Wahrung der Patientenautonomie gemäss den Grundprinzipien der Medizinethik und führt zu mehr Verantwortung des Patienten selbst. Wenn eine Patientin somit hinreichend informiert ist, um selbstständig Nutzen und Risiken abzuwägen, wird sie als Nutzerin eines Systems teilweise verantwortlich. Sollte ein Patient jedoch unzureichend informiert sein oder die Tragweite einer Entscheidung nicht nachvollziehen können, ist es schwierig, die Verantwortung auf diesen zu übertragen. Zudem bleibt zu klären, wie tiefgreifend das Verständnis des KI-Systems sein muss, um die Verantwortung auf den Patienten übertragen zu können.

Die Entwicklerinnen und Entwickler haben sich bei der Frage nach der Verantwortung nie selbst als Verantwortliche bezeichnet. Lediglich auf die Frage, was für ein Gefühl ein Versagen ihres Systems auslösen würde, antworteten die meisten damit, dass dies ein schreckliches Gefühl verursachen und somit zu Schuldgefühlen führen würde. Die Verantwortung, ethische Entscheidungen zu treffen und die Weichen zu stellen, wird von den Entwicklerinnen und Entwicklern mehrheitlich zurückgewiesen. Sie zeigen bei der Frage nach der Verantwortung bei ethischen Fragen eine eher zurückhaltende Position. Im ethischen Entscheidungsprozess wollen sie keine zentrale Rolle übernehmen, auch wenn sie ethische Fragen als etwas Allgegenwärtiges wahrnehmen.

Wenn jemand haftet und verantwortlich gemacht wird, ist es im Moment der Mensch. Diese klare Grenze wird jedoch zunehmend unschärfer, je nachdem, in welche Richtung sich die Rolle der Ärztin oder des Arztes entwickelt und ob und inwieweit KI Bewusstsein und Gefühle entwickeln wird. Bleibt in Zukunft der Arzt und die Ärztin die letzte Instanz?

4.4 Der Wandel des Arztberufs und des Gesundheitswesens durch KI

Künstliche Intelligenz wird uns ersetzen und die Menschheit eliminieren - wird dieses Szenario, wie es im Terminator-Film geschieht, nicht Science-Fiction bleiben, sondern irgendwann Realität? Wird KI Ärztinnen und Ärzte in Zukunft eliminieren? Durch den Fortschritt von KI-Systemen wird sich zumindest das Berufsbild verändern. Ob KI-Anwendungen für den medizinischen Berufsstand eine Chance oder eine Gefahr darstellen, wird sich zeigen.

Eine grosse Studie untersuchte in einer Expertenumfrage die Einschätzung des Zeitpunkts, zu dem KI-Systeme menschliche Leistungen in gewissen Aufgaben erreichen oder übertreffen könnten. Es wurden 2778 Forscherinnen und Forscher, die bei einer der zwei führenden Fachkonferenzen im Bereich Maschinelles Lernen(ICML und NeurIPS) Forschungsarbeiten eingereicht haben, befragt. In der Studie wurden verschiedene menschliche Tätigkeiten untersucht (z.B. «KI setzt LEGO zusammen» oder «Automatisierung aller Berufe»). Die Expertinnen und Experten schätzen den Zeitpunkt, an dem KI zu einer 50% Wahrscheinlichkeit in der Lage sein wird, besser als ein menschlicher Chirurg zu operieren, auf das Jahr 2060.

Einige Ergebnisse der Studie, wie die Forschungsgemeinschaft den Fortschritt der KI in konkreten Anwendungsfeldern einschätzt, sind in *Abbildung* 3 illustriert. Die Abbildung zeigt, wann KI die Leistung eines menschlichen Experten, einer menschlichen Expertin, oder Berufs erreichen oder übertreffen wird.[44]

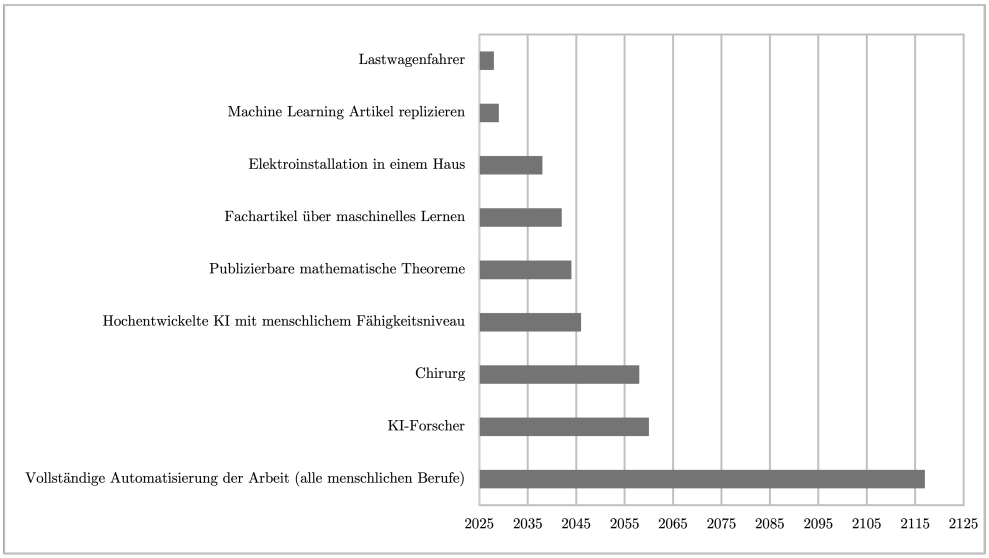

Abbildung 3: Zeitleiste der Schätzung für das Erreichen menschlicher Leistung durch KI

In Anlehnung an Grace, Stewart, Sandkühler, Thomas, Weinstein-Raun und Brauner [44]

Viele der Entwicklerinnen und Entwickler waren in den Interviews dieser Arbeit der Meinung, dass eine Maschine einen Arzt oder eine Ärztin nicht ersetzen könne und sehen die Zukunft des Arztberufs mehrheitlich optimistisch. Sie sind der Ansicht, dass sich der Alltag von Ärztinnen und Ärzten durch die Anwendung von KI verändern, vielleicht verbessern wird und dass sich auf jeden Fall viele Chancen bieten.
Im beruflichen Alltag eines Arztes oder einer Ärztin sollte sich einiges ändern. Bürokratie und Dokumentationspflicht nehmen viel Zeit einer Assistenzärztin oder eines Assistenzarztes in Anspruch. Plötzlich findet man sich als Patientenmanager und nicht mehr als Ärztin wieder. Berichte sollen möglichst schnell und ausführlich geschrieben werden. Sie sind so ausführlich, dass sie oft viele Seiten umfassen und niemand mehr den Überblick über die tatsächlichen Probleme eines Patienten hat. IT-Systeme scheinen aus dem letzten Jahrhundert zu sein. Sie fressen viel Arbeitszeit und generieren teure Arbeitsstunden. Für das Meiste gibt es separate Programme und einzelne Informationen der Patienten müssen mühsam zusammengetragen werden. Es gibt keine einheitliche Patientenakte mehr, in der man auf einen Blick vieles findet. Hätte KI hier nicht ein riesiges Potenzial, um bei dieser Informationsflut und Dokumentationspflicht Hilfe zu leisten und gar bessere Arbeit zu vollbringen?

Einige Entwicklerinnen und Entwickler sind der Meinung, dass KI in der Medizin als Hilfsmittel, also als Werkzeug zum Einsatz kommen wird und soll. Eine KI soll eine Art Unterstützung sein, um die sogenannten «dirty tasks», also mühsame und

zeitraubende Tätigkeiten zu übernehmen. Gerade wenn es um den Aspekt geht, Arbeiten zu übernehmen, die wenig mit dem Arztberuf direkt zu tun haben, sind sich die Befragten einig. Die KI soll eine Unterstützung sein und die Arbeit verbessern und erleichtern. KI in der Medizin könnte durch die Verringerung von «dirty tasks» dazu beitragen, wieder mehr Menschlichkeit in die Medizin zu bringen, wie z.B. auch Haug und Drazen [45] schreiben:

> *We firmly believe that the introduction of AI and machine learning in medicine has helped health professionals improve the quality of care that they can deliver and has the promise to improve it even more in the near future and beyond. Just as computer acquisition of radiographic images did away with the x-ray file room and lost images, AI and machine learning can transform medicine. Health professionals will figure out how to work with AI and machine learning as we grow along with the technology. AI and machine learning will not put health professionals out of business; rather, they will make it possible for health professionals to do their jobs better and leave time for the human–human interactions that make medicine the rewarding profession we all value.*

Eine Entscheidung zu fällen erfordert das Abwägen verschiedener Faktoren gegeneinander. Dabei werden Pro- und Kontra-Argumente verglichen und letztendlich entscheidet man sich für eine Lösung. Auch in der Medizin ist die Entscheidungsfindung nicht anders: Verschiedene Therapieansätze weisen unterschiedliche Vor- und Nachteile auf. Medikamente variieren in ihrem Wirkungs- und Nebenwirkungsprofil. In einem der Interviews wurde beschrieben, dass eine medizinische KI wie ein hilfreicher Berater - wie der Engel auf der Schulter - agieren soll, der Vorschläge und Empfehlungen für die Patientenbehandlung bietet. Sie soll bei der Auswahl der optimalen Therapie für einen Patienten unterstützen und den geeigneten Behandlungsweg aufzeigen. Eine medizinische KI soll helfen, die beste Behandlung für einen Patienten zu finden und den richtigen Pfad zu wählen.

Die Ärzteschaft wird nicht durch KI ersetzt, sondern durch die KI unterstützt werden, sind sich alle Befragten einig. Einer KI würden die menschlichen Eigenschaften fehlen, die zentral in der Arzt-Patienten-Beziehung sind. Empathie könne nicht durch blosse Nachahmung kompensiert werden, diese müsse echt sein, sonst fühle man sich betrogen, lautete eine Begründung. Ein Pflegeroboter, der Empathie imitiert, wäre also nicht gleichwertig wie eine echte Pflegefachperson. Diese imitierte Empathie könnte zwar zu einem gewissen Grad in eine Ärzte-KI einprogrammiert werden, wäre aber dennoch nur ein ablaufendes Programm und keine echte Empathie. KI-Chatbots antworten zwar empathisch, wenn man ihnen erzählt, dass man krank sei, aber dennoch lösen sie nicht das gleiche Gefühl aus, wie wenn einem ein leibhaftiger Mensch gute Besserung wünscht. Das Fehlen von echtem Mitgefühl sei der Hauptunterschied zwischen einer KI und einem Arzt oder einer Ärztin.

Dazu schreiben Weidener und Fischer [46]:

> *It is imperative to ensure that the pursuit of efficiency through AI does not lead to the depersonalization of patient care. Empathy remains a crucial aspect of health care, and AI systems should be used to enhance, rather than replace, the human elements of patient interaction and care.*

Zudem wird hervorgehoben, dass KI-Ethik ein essenzieller Teil in der Ausbildung und in der Lehre im Gesundheitswesen sein soll, um die Menschlichkeit in der Medizin zu bewahren.

Die Expertinnen und Experten bezogen sich in den Interviews fast immer auf einen generellen Mediziner oder eine generelle Medizinerin. Bei Fachbereichen, in denen die direkte Patienteninteraktion weniger relevant ist, wie zum Beispiel in der Radiologie oder Pathologie, wird die Übernahme einer KI als viel wahrscheinlicher eingeschätzt. Das Ersetzen eines Radiologen oder einer Pathologin schien akzeptierter zu sein als in Bereichen, in denen Empathie und Zwischenmenschlichkeit zentrale Aspekte der Behandlung sind.

Eric Topol ist der Meinung, dass Fachbereiche, die stark mit Mustererkennung arbeiten (Radiologie, Dermatologie, Pathologie, Ophthalmologie etc.) sich am stärksten verändern oder gar aussterben werden. Ärzte, wie zum Beispiel Hausärztinnen und allgemeine Mediziner, arbeiten nah am Patienten; in diesen Bereichen geht es um das Zusammenführen von vielen einzelnen Informationen des Patienten (Labor, Bilddaten, Beruf, Sozialleben, etc.) und das Erstellen einer Diagnose und eines passenden Behandlungsplans und natürlich um die Interaktion mit dem Patienten. Er ist der Meinung, dass KI diese nicht ersetzen wird, sondern als Entlastung dienen kann, insbesondere in Bereichen, in denen Maschinen effizienter sind.[17]

In *Abbildung* 3 ist von allen medizinischen Berufen nur der Beruf des Chirurgen und der Chirurgin aufgelistet. Es wurde nicht auf weitere ärztliche Fachbereiche eingegangen. Sie zeigt jedoch, dass gewisse medizinische Fachbereiche durchaus bedroht sind, von KI übernommen zu werden.

Der Konsens ist, dass KI in den nächsten Jahren und Jahrzehnten den Arzt und die Ärztin, zumindest in den meisten Bereichen, nicht ersetzen wird. Für gute Medizin brauche es eine holistische Herangehensweise, die viel mehr als gute Daten und Wissen erfordere. Die Entwickler und Entwicklerinnen sind alle der Meinung, dass KI ein ergänzendes Werkzeug bleiben und den Arzt und die Ärztin als solche nicht ersetzen wird. Der menschliche Faktor bleibe unersetzbar.

5 Schlussfolgerung

Ziel dieser Arbeit war es, sich mithilfe qualitativer Methodik mit ethischen Überlegungen zur Entwicklung, Implementierung und Anwendung von KI-Systemen in der Medizin auseinanderzusetzen und damit die zentrale Forschungsfrage zu beantworten, die den Ausgangspunkt der Untersuchung bildete: Sind sich Expertinnen und Experten aus dem Bereich der medizinischen KI-Entwicklung der Tragweite ihrer Arbeit bewusst und welche ethischen Herausforderungen sehen sie im Moment und in Zukunft? Fragte man die Entwicklerinnen und Entwickler, welche Rolle die Ethik in ihrem Alltag spielt, war die Antwortsuche für viele etwas überfordernd. Wurden jedoch konkretere ethische und moralische Aspekte in den Gesprächen aufgegriffen, ergaben sich viele unterschiedliche ethische Dimensionen, zu denen sich fast alle in ihrer Arbeit Gedanken gemacht hatten. So konnten aus den Interviews vier verschiedene ethische Themengebiete eruiert werden, die in den Resultaten präsentiert und gegliedert wurden:

- Verantwortungsvolle Entwicklung und Implementierung von KI-Systemen
- Sinn und Zweck des Algorithmus
- KI als Werkzeug: Die Zukunft des Arztberufs
- Standards und Richtlinien

Für die Entwicklerinnen und Entwickler prägen viele ethische und moralische Fragen den Alltag: Diese reichen vom Gedanken, ob ihr Programm einen potentiellen Schaden anrichten könnte, über die Hoffnung, dass es Leben retten könnte, bis hin zu den Bedenken, dass ihre Programme auch zur Veränderung des Arztberufs beitragen werden. Aus den ethischen Überlegungen der Entwicklerinnen und Entwickler folgten verschiedene Schwerpunkte, die besonders häufig auftraten oder eine starke Relevanz haben:

- Datenhandhabung und Umgang mit Bias
- Nachvollziehbarkeit und Transparenz: Das Black Box Dilemma
- Verantwortlichkeit und Entscheidung bei der Nutzung von KI in der Medizin
- Der Wandel des Arztberufs und des Gesundheitswesens durch KI

Die Ergebnisse der Arbeit legen nahe, dass Softwareentwicklerinnen und -entwickler zahlreiche ethische und moralische Überlegungen in ihren Berufsalltag integrieren, diese jedoch häufig nicht explizit wahrnehmen. Obwohl sie oft mit ethischen Fragen konfrontiert werden, geschieht dies meist unbewusst. Für fast alle ist die Anwendung von KI in der Medizin mit viel Potenzial und Chancen verbunden. KI könne zu einer besseren, menschlicheren und gerechteren Medizin beitragen, wenn eine ethische Entwicklung und Implementierung gewährleistet werden könne. Eine verantwortungsvolle Entwicklung und Implementierung von KI-Systemen schien allen Befragten wichtig zu sein. Wie ein Algorithmus aufgebaut ist und aus welchen Daten er besteht, soll transparent sein. Es muss klar kommuniziert werden, auf welcher Datenbasis ein System

trainiert wurde und folglich, auf welche Patientengruppe ein KI-System angewendet werden kann. Ein zentraler Punkt dabei ist ein verantwortungsvoller Datenumgang. Die Datensicherheit muss immer gewährleistet werden können, da es sich häufig um sensible Patientendaten handelt. Die Entwicklerinnen und Entwickler sind sich bewusst, dass die Daten, mit denen ein System trainiert wird, ein Knackpunkt für die Güte des KI-Produkts sind. Sowohl die Daten als auch der Algorithmus sind dabei in ihrer Fähigkeit limitiert. Diese Limitation muss offen kommuniziert und transparent sein.

Darf eine medizinische KI eine Black Box sein? Es gibt verschiedene Positionen zur Beantwortung dieser Frage. Entwicklerinnen und Entwickler nehmen häufig eine gewisse Opazität von KI-Systemen in Kauf, wenn das System sicher und gut ist. Anderer Meinung sind die gesetzgebenden und regulierenden Behörden: Eine KI muss so transparent wie möglich sein. KI soll keine Black Box, sondern eine komplett nachvollziehbare «glass box» sein.

Sind die Entwicklerinnen und Entwickler sich der Konsequenzen ihrer Programme ausreichend bewusst? Diese Frage kann nicht gänzlich beantwortet werden. Viele sahen sich nur teilweise in der Verantwortung bei einem Versagen eines Systems. Dennoch ist vielen bewusst, welches Schadens- oder Erfolgspotenzial ihre Systeme haben. Abschliessend kann zur Verantwortlichkeitsfrage gesagt werden, dass ein Diskurs zu diesem Thema notwendig ist. Verantwortung muss möglicherweise auf verschiedene Akteure verteilt werden, die KI kann jedoch bis zum jetzigen Zeitpunkt noch nicht zur Verantwortung gezogen werden und folglich die Verantwortungslücke nicht füllen.

Neue Systeme sollen zu einer Entlastung der Ärztinnen und Ärzte führen, indem sogenannte «dirty tasks» auf KI-Systeme abgewälzt werden können und es so zu einer Effizienzsteigerung und Rehumanisierung des Gesundheitssystems kommen soll. Der Empathie soll wieder mehr Wert beigemessen werden.

Der Wunsch nach ethischer Weiterbildung und mehr Interdisziplinarität wurde angebracht. Es soll eine bessere Zusammenarbeit mit Ethikern oder Ethikerinnen und Ärzten oder Ärztinnen stattfinden. Dies würde eine tiefgreifende Auseinandersetzung mit dem Thema Technologie- und Medizinethik ermöglichen und damit eine verantwortungsvolle sowie nachhaltige Entwicklung von KI-Systemen fördern. Um eine möglichst ethische KI-Entwicklung gewährleisten zu können, soll zwar eine gewisse Regulierung seitens der Behörden stattfinden, jedoch muss gewährleistet sein, dass diese Stellen mit Expertinnen und Experten, die sich sowohl mit den Themen KI, Medizin und Ethik als auch Recht auskennen, besetzt sind. Es zeigte sich, dass der Wunsch nach Regulierung da ist, aber keine Überregulation stattfinden sollte, die den Entwicklungsprozess zu stark bremst. Auch wurde darauf hingewiesen, dass die Ethikkommission unbedingt ein aktuelles und breites Wissen zu Künstlicher Intelligenz mitbringen sollte, damit neue Forschungsarbeiten zugelassen werden können. KI wird den Arztberuf vermutlich in naher Zukunft nicht ersetzen, sondern eher als Werkzeug dienen.

Zum Thema Technologieethik, auch im Kontext der Medizin, wurden bereits einige, vor allem philosophische Abhandlungen veröffentlicht. Dieses Thema soll in der Ausbildung von Medizinern und Softwareentwicklerinnen, Informatikern und Ingenieurinnen Platz finden, damit Medizin ethisch bleibt.

Limitationen dieser Arbeit resultieren aus der Methodik und dem Forschungsdesign.

Die begrenzte Stichprobengrösse bei einer Anzahl von sieben Expertinnen und Experten führt dazu, dass die Umfrage nicht repräsentativ ist. Die Aussagen beruhen auf subjektiven Wahrnehmungen und geben damit wertvolle Einblicke in das Thema, jedoch können auf dieser Grundlage keine allgemein gültigen Aussagen getroffen werden. Durch das Führen von persönlichen Interviews können potenzielle Bias durch die Art der Fragen oder das Setting entstanden sein. Es ist möglich, dass einige Antworten durch soziale Erwünschtheit beeinflusst wurden. Diese Arbeit wurde über den Zeitraum von zwei Jahren geschrieben und da das Feld der KI sehr dynamisch ist, können sich Meinungen in dieser Zeit wieder geändert haben und die Aktualität bei gewissen Aussagen beschränkt sein.

Die Implikation dieser qualitativen Arbeit ist, dass eine ethisch reflektierte und verantwortungsvolle Entwicklung von KI in der Medizin dann erfolgreich sein kann, wenn hochwertige Daten verwendet, Bias erkannt und adressiert sowie die Grenzen und Möglichkeiten der Systeme transparent kommuniziert werden. Das setzt einen interdisziplinären Austausch zwischen Entwicklerinnen, Ärzten und Ethikerinnen voraus. Darüber hinaus sind klare Richtlinien und Gesetze notwendig, um eine sichere Implementierung der Systeme zu gewährleisten. Eine adäquate Regulierung ist dabei wichtig. KI wird helfen, das Gesundheitswesen zu verbessern, womöglich günstiger und wieder menschlicher zu machen, sofern die Ethik darin ihren Platz findet.

Danksagung

Zuallererst gilt mein Dank meinem Betreuer Prof. Dr. Rouven Porz. Mit viel Geduld, fachlichem Rat und moralischer Unterstützung hat er mich durch diese Arbeit begleitet. Ebenso danke ich meinem Freund Florian, mit dem ich unzählige Stunden über ethische Fragen und Künstliche Intelligenz diskutieren durfte. Ein grosser Dank gilt auch meiner Mutter Conny, die mir zugehört hat und mich motivierte, wenn ich nicht mehr weiter wusste. Nicht zuletzt danke ich meinen engen und treuen Freunden, die mir mit ihrer Unterstützung und ihrem Rückhalt jederzeit zur Seite stehen.

Literaturverzeichnis

1. Shelley M. Frankenstein oder Der moderne Prometheus. Reclam.
2. Schwab G. Sagen des klassischen Altertums. Buch von Gustav Schwab. 3. Aufl. Insel Verlag, 2017.
3. Nyholm S. Humans and robots: : ethics, agency, and anthropomorphism. Unter Mitarb. von Lyndon B Johnson Space Center. Philosophy, Technology and Society. London ; Rowman & Littlefield International, Ltd., 2020.
4. Kunzmann P, Burkard FP und Wiedmann F. dtv-Atlas zur Philosophie: Tafeln und Texte. Deutscher Taschenbuch Verlag, 1991. 268 S.
5. Prechtl P und Burkard FP, Hrsg. Metzler Lexikon Philosophie. Stuttgart: J.B. Metzler, 2008. (Besucht am 13. 03. 2025).
6. Beauchamp T und Childress J. Principles of Biomedical Ethics: Marking Its Fortieth Anniversary. The American journal of bioethics: AJOB 2019;19:9–12.
7. Egger M, Razum O und Rieder A. Public Health Kompakt. Hrsg. von Egger M, Razum O und Rieder A. Unter Mitarb. von Egger M, Razum O und Rieder A. Berlin: de Gruyter, 2018. 525 S. (Besucht am 23. 09. 2024).
8. Nyholm S. This is Technology Ethics: An Introduction. 1. Aufl. Wiley-Blackwell, 2022. (Besucht am 02. 11. 2023).
9. EU AI Act: first regulation on artificial intelligence. Topics | European Parliament. 2023. (Besucht am 24. 09. 2024).
10. Fulterer R. Fachleute aus der ganzen Welt testen in St. Gallen KI-Regeln. Neue Zürcher Zeitung 2023.
11. Ein nüchterner Blick auf den Mythos KI: Auswirkungen von Künstlicher Intelligenz auf den Menschen. 2024. (Besucht am 24. 09. 2024).
12. Forderungen an die Künstliche Intelligenz (KI) aus Sicht der FMH. 2022. (Besucht am 24. 09. 2024).
13. McCarthy J, Minsky ML, Rochester N und Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine 2006;27:12–2.
14. Bringsjord S und Govindarajulu NS. Artificial Intelligence. In: *The Stanford Encyclopedia of Philosophy*. Hrsg. von Zalta EN und Nodelman U. Fall 2024. Metaphysics Research Lab, Stanford University, 2024. (Besucht am 27. 09. 2024).
15. Wichert A. Künstliche Intelligenz. (Besucht am 30. 09. 2024).
16. Krauss P. Künstliche Intelligenz und Hirnforschung: Neuronale Netze, Deep Learning und die Zukunft der Kognition. Berlin, Heidelberg: Springer, 2023. (Besucht am 20. 01. 2025).
17. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, 2019. 373 S.

18. Goodfellow I, Bengio Y und Courville A. Deep Learning. The MIT Press, 2016. 800 S.
19. Mainzer K und Kahle R. Grenzen der KI – theoretisch, praktisch, ethisch. Technik im Fokus. Berlin, Heidelberg: Springer, 2022. (Besucht am 05. 01. 2025).
20. Campbell M, Hoane AJ und Hsu Fh. Deep Blue. Artificial Intelligence 2002;134:57–83.
21. Schwartz WB. Medicine and the Computer: The Promise and Problems of Change. In: *Use and Impact of Computers in Clinical Medicine*. Hrsg. von Anderson JG und Jay SJ. New York, NY: Springer, 1987:321–35. (Besucht am 02. 12. 2024).
22. Jutzeler C. KI in der Medizin: Schlüssel für personalisierte Therapien. Neue Zürcher Zeitung 2024.
23. FMH. Standesordnung der FMH.
24. Flick U, Kardorff Ev und Steinke I. Qualitative Forschung: ein Handbuch. 14. Auflage, Originalausgabe. Rororo 55628 ; Rowohlts Enzyklopädie. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag, 2022.
25. Pietkiewicz I und Smith JA. A practical guide to using interpretative phenomenological analysis in qualitative research psychology. Psychological journal 2014;20:7–14.
26. Porz R. Zwischen Entscheidung und Entfremdung: Patientenperspektiven in der Gendiagnostik und Albert Camus' Konzepte zum Absurden : eine empirisch-ethische Interviewstudie. Diss. Paderborn: Mentis, 2008.
27. Moher D, Altman DG, Liberati A, Tetzlaff J, Takkouche B und Norman G. PRISMA Statement. Epidemiology (Cambridge, Mass.) 2011;22:128–8.
28. Rawls J. A Theory of Justice. Harvard University Press, 1971. (Besucht am 17. 01. 2025).
29. Bostrom N und Bostrom N. Superintelligence: paths, dangers, strategies. First edition. Oxford: University Press, 2014. 328 S.
30. AI Value Alignment for Shared Human Goals. World Economic Forum. (Besucht am 03. 04. 2025).
31. Obermeyer Z, Powers B, Vogeli C und Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–53.
32. Zimmer A. Künstliche Intelligenz im ärztlichen Alltag. 2022.
33. Rajkomar A, Hardt M, Howell MD, Corrado G und Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Annals of internal medicine 2018;169:866–72.
34. Swissmedic 2019 ©C. Regulierung Medizinprodukte. (Besucht am 17. 02. 2025).
35. Stahl BC. Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. SpringerBriefs in Research and Innovation Governance. Cham: Springer International Publishing, 2021. (Besucht am 21. 09. 2024).

36. Knight W. The Dark Secret at the Heart of AI. MIT Technology Review. 2017. (Besucht am 17. 01. 2025).

37. Wynsberghe A van. Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics 2021;1:213–8.

38. Henkel CH. Wenn der «Roboter am Lenkrad» einen Fehler macht. Neue Zürcher Zeitung 2016.

39. ChatGPT. (Besucht am 23. 01. 2025).

40. Santoni de Sio F und Mecacci G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. Philosophy & Technology 2021;34:1057–84.

41. Datenschutz-Grundverordnung (DSGVO) | EUR-Lex. (Besucht am 28. 01. 2025).

42. Beckers A und Teubner G. Die digitale Verantwortungslücke: Vorschläge zur Haftung für algorithmisches Fehlverhalten. In: *Künstliche Intelligenz, Mensch und Gesellschaft: Soziale Dynamiken und gesellschaftliche Folgen einer technologischen Innovation*. Hrsg. von Heinlein M und Huchler N. Wiesbaden: Springer Fachmedien, 2024:153–77. (Besucht am 27. 01. 2025).

43. South Park - Deep Learning | South Park Studios Deutsch. 2023. (Besucht am 16. 11. 2023).

44. Grace K, Stewart H, Sandkühler JF, Thomas S, Weinstein-Raun B und Brauner J. Thousands of AI Authors on the Future of AI. 2024. arXiv: `2401.02843[cs]`. (Besucht am 02. 05. 2025).

45. Haug CJ und Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. New England Journal of Medicine 2023;388:1201–8.

46. Weidener L und Fischer M. Proposing a Principle-Based Approach for Teaching AI Ethics in Medical Education. JMIR Medical Education 2024;10:e55368.

Anhang

Endversion des Leitfadens

Experteninterview – Ethische Schwerpunkte im Zusammenhang mit der Entwicklung von Künstlicher Intelligenz in der Medizin

Maria Mahler, cand. Dr. med., Universität Bern
Prof. Dr. phil. Rouven Porz, Medizinethik, Inselspital Bern

Informationen

Das folgende Interview soll ungefähr 30 Minuten dauern. Es handelt sich um ein qualitatives, semistrukturiertes Interview, das in fünf Teile gegliedert ist, mit Fokus auf Teil drei und Teil vier. Die Fragen werden so offen wie möglich formuliert. Je nach Antwort können auch Anschlussfragen gestellt oder spätere Fragen entfallen.

Forschungsausrichtung

Ich sehe Sie als Expertin oder Experten auf dem Gebiet der KI-Entwicklung in der Medizin. Daher möchte ich gerne mehr über Ihr Verständnis und Ihre Gedanken zur Ethik in diesem Fachbereich erfahren. Eine quantitative und normative Auswertung der Daten ist nicht beabsichtigt. Vielmehr interessieren mich Ihre persönliche Ansicht und Ihr Wissen.

Datenauswertung

Die Interviews werden mittels qualitativer Forschungsmethoden ausgewertet. Sie werden akustisch aufgezeichnet, transkribiert und anonymisiert. Das Gespräch unterliegt der medizinischen Schweigepflicht.

Offene Fragen

Haben Sie noch Fragen vor Beginn des Interviews?

Einverständniserklärung

Ich erkläre mich hiermit einverstanden mit der Aufzeichnung und der Verwendung meiner anonymisierten Daten für diese Doktorarbeit.

Name, Vorname:

Ort, Datum: Unterschrift:

Interview

Einstieg / Persönliche Vorstellung

Zum Einstieg werde ich Ihnen Fragen zu Ihrer Person und zu Ihrer Arbeit stellen, Sie können diese gerne kurz und knapp beantworten.

- Woran arbeiten Sie zurzeit?
- Wie war Ihr beruflicher Werdegang?
- Was war für Sie ausschlaggebend, sich beruflich mit KI zu befassen?

Persönlicher Bezug zu KI

Die folgenden Fragen werden etwas offener sein und beziehen sich auf Ihren persönlichen Bezug zu KI.

- Wann sind Sie das erste Mal mit dem Begriff KI in Berührung gekommen? Was hat das bei ihnen ausgelöst?
- Wie würden Sie KI beschreiben? Haben Sie eine eigene Definition dafür?
- Gibt es ein KI-Projekt in der Medizin, das Sie zurzeit besonders spannend finden (eigenes oder anderes)?

Fragen im Hinblick auf Ethik im Bereich KI

Die folgenden Fragen befassen sich mit dem Thema Ethik im Bereich der KI-Entwicklung. Diese sind etwas genereller gestellt und sie können gerne auch etwas ausholen.

- Was verstehen sie unter Ethik?
- Arbeiten Sie mit externen/internen Fachpersonen im Bereich Ethik zusammen?
- Welche ethischen Überlegungen spielen im Berufsalltag eine Rolle? Können Sie mir einen Einblick in Ihre Gedanken geben?

A. Verantwortung / Entscheidung

B. Datensicherheit

C. Bias

D. Transparenz

E. Horizont / Limitation / Zukunft

F. Kosten

G. Selektion / Personalisierte Medizin

H. Mensch vs. Technik

I. Nachhaltigkeit

A) Verantwortung / Entscheidung

- Kann KI Verantwortung übernehmen?
- Wer trägt Ihrer Meinung nach die Verantwortung bei der Entwicklung und Anwendung von KI-Technologien? Sehen Sie gewisse Verantwortungslücken?
- Sollte die letzte Entscheidung immer beim Menschen/Arzt liegen? Oder macht es Sinn, diese Entscheidung an KI/Technologie abzugeben? (z.B. Operationsroboter)

B) Datensicherheit

- Sehen Sie generell ein Problem mit der Datensicherheit im KI-Bereich?
- Würden Sie ihre Gesundheitsdaten bei einem KI-Diagnostikprogramm einspeisen? Warum Ja/Nein?
- Wie stehen Sie zum Einsatz von grossen Datenbanken wie z.B. BioBank aus England, mit Daten von über 500'000 Personen?

C) Bias

- Wie kann sichergestellt werden, dass gesellschaftliche Ungerechtigkeiten (Diskriminierung, Rassismus, Armut etc.) sich nicht in einer KI widerspiegeln?
- Bräuchte es dafür eine zusätzliche moralische Instanz, um diese Daten von Ungerechtigkeit zu befreien?

D) Transparenz

- Denken Sie, dass KI eine Black Box ist?
- Ist es ethisch vertretbar, eine KI in der Medizin zu verwenden, deren Algorithmus und Entscheidungen nicht mehr nachvollziehbar sind?

E) Horizont / Limitation / Zukunft

- Mit dem Einzug von KI in immer mehr Bereiche des Lebens, steigt der Energieverbrauch zur Verarbeitung der Daten exponentiell an. Ist es denkbar, dass die Entwicklung von KI-Technologien nicht zuerst an technische oder politische, sondern an Energie / Ressourcen bedingte Grenzen stösst.

F) Kosten / Gesundheitskosten

- Die steigenden Gesundheitskosten und den damit verbundenen stetig steigenden Krankenkassenprämien sind ein zunehmendes Problem. Wo sehen Sie die Rolle der KI in dieser Entwicklung?

- Wäre es sinnvoll im Rahmen der steigenden Gesundheitskosten, individuelle Krankenkassen- und Versicherungsprämien mittels der erhobenen Gesundheitsdaten gespeister KI zu bestimmen?

G) Selektion / Personalisierte Medizin

- Bestimmt sind Sie bekannt mit Projekten der personalisierten Medizin wie «Deep genomics»: (DNA / Genom Analyse / Krankheitserkennung und individueller Entwicklung der Therapie mittels KI anhand des Genoms) -> Wo sehen Sie das Potenzial und wo die Gefahren von KI basierter Genomanalyse?

H) Technik vs. Mensch

- Würden Sie lieber von einem Arzt mit einer hohen Fehlerquote behandelt werden oder von einer Diagnostik-KI mit einer niedrigen Fehlerquote.

Ethik im Kontext KI in der Medizin

Als nächstes folgen einige Fragen spezifisch zur Ethik im Bereich KI in der Medizin.

- Was empfinden Sie, wenn eine KI-Technologie, zu deren Entwicklung sie beigetragen haben erfolgreich ist und z.B. eine richtige Diagnose liefert?
- Was empfinden Sie oder würden Sie empfinden, wenn eine KI-Technologie, die Sie mitentwickelt haben, nicht wie gewünscht arbeitet oder allenfalls zu einer falschen Diagnose / Behandlung oder Intervention führen würde?
- Welche Qualitäten machen einen guten Arzt oder eine gute Ärztin aus?
- Wäre es möglich diese Qualitäten einer KI zu geben?

Mögliche Stichworte:

A. Verantwortung / Entscheidung

B. Datensicherheit

C. Bias

D. Transparenz

E. Horizont / Limitation / Zukunft

F. Kosten

G. Selektion / Personalisierte Medizin

H. Mensch vs. Technik

I. Nachhaltigkeit

Abschlussfragen

Jetzt kommen wir zum letzten Teil des Interviews.

- Ist der Arztberuf, wie wir ihn kennen bereits auf der Palliativstation??
- Welche Ratschläge würden Sie einer angehenden Ärztin im Umgang mit KI in der Medizin mitgeben?

Ergänzende Fragen (Diese Fragen können im Anschluss noch gestellt werden, falls genügend Zeit bleibt und bilden den Abschluss des Gesprächs.)

- Stichwort Cyborg: Wie lange ist ein Mensch noch ein Mensch und wann hört das Geschöpf auf Mensch zu sein? Wenn man nach und nach alle Teile des Menschen durch künstliche Komponenten ersetzen würde: Ist es ein fliessender Übergang oder gibt es ein «Menschteil» und wenn dieses Entfernt wird, ist der Mensch im Menschen weg.
- Vorausgesetzt eine KI hätte die Aufgabe, möglichst viele Leben zu retten, wieso sollte diese dann nicht die Organe eines Menschen (z.B. eines Mörders) entnehmen (in dem Vorgang wird dieser getötet), um 5 andere zu retten?